

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/31, C07K 14/22, 16/12, A61K 39/095 // (C12N 15/31, C12R 1:36)		A2	(11) International Publication Number: WO 99/24578
			(43) International Publication Date: 20 May 1999 (20.05.99)
(21) International Application Number: PCT/IB98/01665		(74) Agent: HALLYBONE, Huw, George; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).	
(22) International Filing Date: 9 October 1998 (09.10.98)			
(30) Priority Data:		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
9723516.2 6 November 1997 (06.11.97) GB			
9724190.5 14 November 1997 (14.11.97) GB			
9724386.9 18 November 1997 (18.11.97) GB			
9725158.1 27 November 1997 (27.11.97) GB			
9726147.3 10 December 1997 (10.12.97) GB			
9800759.4 14 January 1998 (14.01.98) GB			
9819016.8 1 September 1998 (01.09.98) GB			
(71) Applicant (for all designated States except US): CHIRON S.P.A. [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT).		Published Without international search report and to be republished upon receipt of that report.	
(72) Inventors; and			
(75) Inventors/Applicants (for US only): MASIGNANI, Vega [IT/IT]; Via Pantaneto, 105, I-53100 Siena (IT). RAP-PUOLI, Rino [IT/IT]; Via delle Rocche, 1, Vagliagli, I-53019 Castelnuovo Berardenga (IT). PIZZA, Mariagrazia [IT/IT]; Strada di Montalbuccio, 160, I-53100 Siena (IT). SCARLATO, Vincenzo [IT/IT]; Via Firenze, 3/37, I-53134 Colle Val d'Elsa (IT). GRANDI, Guido [IT/IT]; 9° Strada, 4, I-20090 Segrate (IT).			
(54) Title: NEISSERIAL ANTIGENS			
(57) Abstract			
The invention provides proteins from <i>Neisseria meningitidis</i> (strains A and B) and from <i>Neisseria gonorrhoeae</i> including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LJ	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

NEISSERIAL ANTIGENS

This invention relates to antigens from *Neisseria* bacteria.

BACKGROUND ART

Neisseria meningitidis and *Neisseria gonorrhoeae* are non-motile, gram negative diplococci that are pathogenic in humans. *N.meningitidis* colonises the pharynx and causes meningitis (and, occasionally, septicaemia in the absence of meningitis); *N.gonorrhoeae* colonises the genital tract and causes gonorrhea. Although colonising different areas of the body and causing completely different diseases, the two pathogens are closely related, although one feature that clearly differentiates meningococcus from gonococcus is the presence of a polysaccharide capsule that is present in all pathogenic meningococci.

N.gonorrhoeae caused approximately 800,000 cases per year during the period 1983-1990 in the United States alone (chapter by Meitzner & Cohen, "Vaccines Against Gonococcal Infection", In: *New Generation Vaccines*, 2nd edition, ed. Levine, Woodrow, Kaper, & Cobon, Marcel Dekker, New York, 1997, pp.817-842). The disease causes significant morbidity but limited mortality. Vaccination against *N.gonorrhoeae* would be highly desirable, but repeated attempts have failed. The main candidate antigens for this vaccine are surface-exposed proteins such as pili, porins, opacity-associated proteins (Opas) and other surface-exposed proteins such as the Lip, Laz, IgA1 protease and transferrin-binding proteins. The lipooligosaccharide (LOS) has also been suggested as vaccine (Meitzner & Cohen, *supra*).

N.meningitidis causes both endemic and epidemic disease. In the United States the attack rate is 0.6-1 per 100,000 persons per year, and it can be much greater during outbreaks (see Lieberman *et al.* (1996) Safety and Immunogenicity of a Serogroups A/C *Neisseria meningitidis* Oligosaccharide-Protein Conjugate Vaccine in Young Children. *JAMA* 275(19):1499-1503; Schuchat *et al* (1997) Bacterial Meningitis in the United States in 1995. *N Engl J Med* 337(14):970-976). In developing countries, endemic disease rates are much higher and during epidemics incidence rates can reach 500 cases per 100,000 persons per year. Mortality is extremely high, at 10-20% in the United States, and much higher in developing countries. Following the introduction of the conjugate vaccine against *Haemophilus influenzae*, *N. meningitidis* is the major cause of bacterial meningitis at all ages in the United States (Schuchat *et al* (1997) *supra*).

Based on the organism's capsular polysaccharide, 12 serogroups of *N.meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for the rest of the cases in the United States and developed countries. The meningococcal vaccine currently in use is a tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Although efficacious in adolescents and adults, it induces a poor immune response and short duration of protection, and cannot be used in infants [eg. Morbidity and Mortality weekly report, Vol.46, No. RR-5 (1997)]. This is because polysaccharides are T-cell independent antigens that induce a weak immune response that cannot be boosted by repeated immunization. Following the success of the vaccination against *H.influenzae*, conjugate vaccines against serogroups A and C have been developed and are at the final stage of clinical testing (Zollinger WD "New and Improved Vaccines Against Meningococcal Disease" in: *New Generation Vaccines*, supra, pp. 469-488; Lieberman *et al* (1996) supra; Costantino *et al* (1992) Development and phase I clinical testing of a conjugate vaccine against meningococcus A and C. *Vaccine* 10:691-698).

Meningococcus B remains a problem, however. This serotype currently is responsible for approximately 50% of total meningitis in the United States, Europe, and South America. The polysaccharide approach cannot be used because the menB capsular polysaccharide is a polymer of $\alpha(2-8)$ -linked *N*-acetyl neuraminic acid that is also present in mammalian tissue. This results in tolerance to the antigen; indeed, if an immune response were elicited, it would be anti-self, and therefore undesirable. In order to avoid induction of autoimmunity and to induce a protective immune response, the capsular polysaccharide has, for instance, been chemically modified substituting the *N*-acetyl groups with *N*-propionyl groups, leaving the specific antigenicity unaltered (Romero & Outschoorn (1994) Current status of Meningococcal group B vaccine candidates: capsular or non-capsular? *Clin Microbiol Rev* 7(4):559-575).

Alternative approaches to menB vaccines have used complex mixtures of outer membrane proteins (OMPs), containing either the OMPs alone, or OMPs enriched in porins, or deleted of the class 4 OMPs that are believed to induce antibodies that block bactericidal activity. This approach produces vaccines that are not well characterized. They are able to protect against the homologous strain, but are not effective at large where there are many antigenic variants of the outer membrane proteins. To overcome the antigenic variability, multivalent vaccines containing up to nine different

porins have been constructed (eg. Poolman JT (1992) Development of a meningococcal vaccine. *Infect. Agents Dis.* 4:13-28). Additional proteins to be used in outer membrane vaccines have been the opa and opc proteins, but none of these approaches have been able to overcome the antigenic variability (eg. Ala'Aldeen & Borriello (1996) The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* 14(1):49-53).

A certain amount of sequence data is available for meningococcal and gonococcal genes and proteins (eg. EP-A-0467714, WO96/29412), but this is by no means complete. The provision of further sequences could provide an opportunity to identify secreted or surface-exposed proteins that are presumed targets for the immune system and which are not antigenically variable. For instance, some of the identified proteins could be components of efficacious vaccines against meningococcus B, some could be components of vaccines against all meningococcal serotypes, and others could be components of vaccines against all pathogenic *Neisseriae*.

THE INVENTION

The invention provides proteins comprising the Neisserial amino acid sequences disclosed in the examples. These sequences relate to *N.meningitidis* or *N.gonorrhoeae*.

It also provides proteins comprising sequences homologous (*ie.* having sequence identity) to the Neisserial amino acid sequences disclosed in the examples. Depending on the particular sequence, the degree of identity is preferably greater than 50% (*eg.* 65%, 80%, 90%, or more). These homologous proteins include mutants and allelic variants of the sequences disclosed in the examples. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between the proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

The invention further provides proteins comprising fragments of the Neisserial amino acid sequences disclosed in the examples. The fragments should comprise at least *n* consecutive amino acids from the sequences and, depending on the particular sequence, *n* is 7 or more (*eg.* 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments comprise an epitope from the sequence.

The proteins of the invention can, of course, be prepared by various means (*eg.* recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (*eg.* native, fusions *etc.*). They are preferably prepared in substantially pure or isolated form (*ie.* substantially free from other Neisserial or host cell proteins)

- 5 According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means.

According to a further aspect, the invention provides nucleic acid comprising the Neisserial nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences homologous (*ie.* having sequence identity) to the Neisserial nucleotide
10 sequences disclosed in the examples.

Furthermore, the invention provides nucleic acid which can hybridise to the Neisserial nucleic acid disclosed in the examples, preferably under "high stringency" conditions (*eg.* 65°C in a 0.1xSSC, 0.5% SDS solution).

Nucleic acid comprising fragments of these sequences are also provided. These should comprise
15 at least *n* consecutive nucleotides from the Neisserial sequences and, depending on the particular sequence, *n* is 10 or more (*eg.* 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences
20 complementary to those described above (*eg.* for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*eg.* single stranded, double stranded, vectors, probes *etc.*).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as
25 those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (eg. expression vectors) and host cells transformed with such vectors.

According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as vaccines,
5 for instance, or as diagnostic reagents, or as immunogenic compositions.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (eg. as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to Neisserial bacteria; (ii) a diagnostic reagent for detecting the
10 presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria; and/or (iii) a reagent which can raise antibodies against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as *N.gonorrhoeae*, or any strain of *N.meningitidis*, such as strain A, strain B or strain C).

The invention also provides a method of treating a patient, comprising administering to the patient
15 a therapeutically effective amount of nucleic acid, protein, and/or antibody according to the invention.

According to further aspects, the invention provides various processes.

A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

20 A process for producing protein or nucleic acid of the invention is provided, wherein the the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting said duplexes.

25 A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

A summary of standard techniques and procedures which may be employed in order to perform the invention (eg. to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

5 General

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature eg. Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and*
10 *ii* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene*
15 *Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

20 Standard abbreviations for nucleotides and amino acids are used in this specification.

All publications, patents, and patent applications cited herein are incorporated in full by reference. In particular, the contents of UK patent applications 9723516.2, 9724190.5, 9724386.9, 9725158.1, 9726147.3, 9800759.4, and 9819016.8 are incorporated herein.

Definitions

25 A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" eg. a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a
5 Neisserial sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.

An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous
10 unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7
15 cells.

A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (eg. 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the
20 Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes
25 a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (eg. see US patent 5,753,235).

Expression systems

The Neisserial nucleotide sequences can be expressed in a variety of different expression systems;
30 for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

i. Mammalian Systems

Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) *Proc. Natl. Acad. Sci.* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only

in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired,
5 the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing
10 sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

15 Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In
20 *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

25 Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as
30 mammalian cells or bacteria. Mammalian replication systems include those derived from animal

- viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].
- 10 The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.
- 15 Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (eg. Hep G2), and a number of other cell lines.
- 20 ii. Baculovirus Systems
- The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.
- 25
- After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques
- 30

are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its owned set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 5 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by 10 insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human α -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also 15 be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence 20 containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised 25 of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer 30 vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter

- and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion
- 5 can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.
- 10 The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein,
- 15 which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 μ m in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from
- 20 wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).
- 25 Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, *et al.* (1989) *In*
- 30 *Vitro Cell. Dev. Biol.* 25:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. *See, eg. Summers and Smith supra.*

5 The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as
10 chromatography, eg. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

15 In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

iii. Plant Systems

20 There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in
25 Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by
30 gibberellic acid can be found in R.L. Jones and J. MacMillin, *Gibberellins: in: Advanced Plant Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52.

References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Repr.* 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl. Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (eg. structural gene) into mRNA. A promoter will have a transcription initiation
5 region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and
10 thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in Escherichia coli (E.
15 coli) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include
20 promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The g-laotamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406]
25 promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac*
30 promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21].

Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* on *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.*

(1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (eg. ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghrayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal

element (eg. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907],

Streptococcus cremoris [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; *Streptococcus lividans* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptomyces lividans* [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl₂ or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See *eg.* [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; *Escherichia*], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 *Lactobacillus*]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, *Pseudomonas*]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, *Staphylococcus*], [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of *Streptococcus lactis* by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology* 1:412, *Streptococcus*].

v. Yeast Expression

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene.

The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence

of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

- Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].
- 10 In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters
- 15 which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].
- 20
- 25 A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (*eg.* see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (eg. plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See eg. Brake *et al.*, *supra*.

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the

chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135], *Pichia guilliermondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*];

[Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; Hansenula]; [Das *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; Kluyveromyces]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patent
5 Nos. 4,837,148 and 4,929,555; Pichia]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 Saccharomyces]; [Beach and Nurse (1981) *Nature* 300:706; Schizosaccharomyces]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; Yarrowia].

Antibodies

10 As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised
15 antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying Neisserial proteins.

Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably
20 a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 µg/injection
25 is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by in vitro immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating
30 the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is

recovered by centrifugation (eg. 1,000g for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described
5 above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of
10 the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (eg. hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then
15 cultured either *in vitro* (eg. in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes
20 are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A,
25 and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example, ^{125}I may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of
30 this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with ^{125}I , or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be

readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Vaccines

Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, etc. pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents

such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalene, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (eg. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (eg. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

The immunogenic compositions (eg. the immunising antigen/immunogen/polypeptide/protein/nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (eg. nonhuman primate, primate, etc.), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, eg. by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (eg. WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [eg. Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

20 Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence *in vivo* can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus,

picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

5 Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses eg. MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

10 Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

15 These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

20 Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (eg. HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

25 Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia, Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or

collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; 5 WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 10 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors 15 employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. 20 Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in 25 which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted 30 terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the

native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470.

10 Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional

15 exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

20 Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and

25 WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN

30 08/679640).

DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

- 5 Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317;
- 10 Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805;
- 15 Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240;
- 20 Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245;
- 25 Tonate virus, for example ATCC VR-925; Trinita virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre
- 30 (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid

- expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.
- 10 Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.*
- 15 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

- Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the
- 20 beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

- Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional
- 25 vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active
- 30 promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA*

91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, Biochemistry, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

15 Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

20 Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (eg. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

25 Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in eg. WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A. Polypeptides

- One example are polypeptides which include, without limitation: asialoglycoprotein (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

B. Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C. Polyalkylenes, Polysaccharides, etc.

Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D. Lipids, and Liposomes

The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

- Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the

use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to
5 mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1,2,3-dioleoyloxy]propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand
10 Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, eg. Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

15 Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate
20 ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See eg. Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta*
25 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

E. Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, and E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, and E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, and E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol.* (*supra*); Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30:

443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

F. Polycationic Agents

- 5 Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both in vitro, ex vivo, and in vivo applications. Polycationic agents can
10 be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful
15 as nucleic acid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the
20 list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

25 Immunodiagnostic Assays

Neisserial antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-Neisserial antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to Neisserial proteins within biological samples, including for example, blood or serum

samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody
5 or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed
10 by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, *etc.*) required for the conduct of the assay, as well as suitable set of assay instructions.

Nucleic Acid Hybridisation

"Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen
15 bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences;
20 use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

"Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt
25 concentration should be chosen that is approximately 120 to 200°C below the calculated T_m of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

30 Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The

- total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to 10^{-9} to 10^{-8} g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8 cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/µg, resulting in an exposure time of ~24 hours.
- Several factors can affect the melting temperature (T_m) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4[\%(G + C)] - 0.6(\% \text{formamide}) - 600/n - 1.5(\% \text{mismatch}).$$

where C_i is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

- In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

- In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology,

and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed
5 after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid
10 probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the Neisserial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will
15 encode the amino acid sequence, the native Neisserial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Neisserial sequence (or its complement) — some
20 variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Neisserial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe
25 sequence being complementary to a Neisserial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Neisserial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as
30 temperature, salt condition and the like. For example, for diagnostic applications, depending on the

complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

- 5 Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.*
10 backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

- 15 Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 and 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with
20 duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Neisserial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern
25 blots. When using the Southern blot method, the labelled probe will hybridize to the Neisserial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid
30 support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed

to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-20 show biochemical data obtained in the Examples, and also sequence analysis, for ORFs 37, 5, 2, 15, 22, 28, 32, 4, 61, 76, 89, 97, 106, 138, 23, 25, 27, 79, 85 and 132. M1 and M2 are molecular weight markers. Arrows indicate the position of the main recombinant product or, in Western blots, the position of the main *N.meningitidis* immunoreactive band. TP indicates *N.meningitidis* total protein extract; OMV indicates *N.meningitidis* outer membrane vesicle preparation. In bactericidal assay results: a diamond (♦) shows preimmune data; a triangle (▲) shows GST control data; a circle (●) shows data with recombinant *N.meningitidis* protein. Computer analyses show a hydrophilicity plot (upper), an antigenic index plot (middle), and an AMPHI analysis (lower). The AMPHI program has been used to predict T-cell epitopes [Gao *et al.* (1989) *J. Immunol.* 143:3007; Roberts *et al.* (1996) *AIDS Res Hum Retrovir* 12:593; Quakyi *et al.* (1992) *Scand J Immunol* suppl.11:9) and is available in the Protean package of DNASTAR, Inc. (1228 South Park Street, Madison, Wisconsin 53715 USA).

EXAMPLES

The examples describe nucleic acid sequences which have been identified in *N.meningitidis*, along with their putative translation products, and also those of *N.gonorrhoeae*. Not all of the nucleic acid sequences are complete *ie.* they encode less than the full-length wild-type protein.

- The examples are generally in the following format:
- a nucleotide sequence which has been identified in *N.meningitidis* (strain B)
 - the putative translation product of this sequence
 - a computer analysis of the translation product based on database comparisons
 - corresponding gene and protein sequences identified in *N.meningitidis* (strain A) and in *N.gonorrhoeae*
 - a description of the characteristics of the proteins which indicates that they might be suitably antigenic
 - results of biochemical analysis (expression, purification, ELISA, FACS *etc.*)

The examples typically include details of sequence identity between species and strains. Proteins that are similar in sequence are generally similar in both structure and function, and the sequence identity often indicates a common evolutionary origin. Comparison with sequences of proteins of known function is widely used as a guide for the assignment of putative protein function to a new sequence and has proved particularly useful in whole-genome analyses.

Sequence comparisons were performed at NCBI (<http://www.ncbi.nlm.nih.gov>) using the algorithms BLAST, BLAST2, BLASTn, BLASTp, tBLASTn, BLASTx, & tBLASTx [eg. see also Altschul *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:2289-3402]. Searches were performed against the following databases: non-redundant GenBank+EMBL+DDBJ+PDB sequences and non-redundant GenBank CDS translations+PDB+SwissProt+SPupdate+PIR sequences.

To compare Meningococcal and Gonococcal sequences, the tBLASTx algorithm was used, as implemented at http://www.genome.ou.edu/gono_blast.html. The FASTA algorithm was also used to compare the ORFs (from GCG Wisconsin Package, version 9.0).

Dots within nucleotide sequences (eg. position 495 in SEQ ID 11) represent nucleotides which have been arbitrarily introduced in order to maintain a reading frame. In the same way, double-underlined nucleotides were removed. Lower case letters (eg. position 496 in SEQ ID 11) represent ambiguities which arose during alignment of independent sequencing reactions (some of the nucleotide sequences in the examples are derived from combining the results of two or more experiments).

Nucleotide sequences were scanned in all six reading frames to predict the presence of hydrophobic domains using an algorithm based on the statistical studies of Esposti *et al.* [Critical evaluation of the hydropathy of membrane proteins (1990) *Eur J Biochem* 190:207-219]. These domains represent potential transmembrane regions or hydrophobic leader sequences.

Open reading frames were predicted from fragmented nucleotide sequences using the program ORFFINDER (NCBI).

Underlined amino acid sequences indicate possible transmembrane domains or leader sequences in the ORFs, as predicted by the PSORT algorithm (<http://www.psорт.nibb.ac.jp>). Functional domains were also predicted using the MOTIFS program (GCG Wisconsin & PROSITE).

Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *ie.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The recombinant protein can also be conveniently used to prepare antibodies *eg.* in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (*eg.* fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

10 In particular, the following methods (A) to (S) were used to express, purify and biochemically characterise the proteins of the invention:

A) Chromosomal DNA preparation

N.meningitidis strain 2996 was grown to exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in 5ml buffer (20% Sucrose, 50mM Tris-HCl, 50mM EDTA, pH8).
15 After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50µg/ml Proteinase K), and the suspension was incubated at 37°C for 2 hours. Two phenol extractions (equilibrated to pH 8) and one ChCl₃/isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes ethanol, and was collected by centrifugation. The pellet was washed once with 70%
20 ethanol and redissolved in 4ml buffer (10mM Tris-HCl, 1mM EDTA, pH 8). The DNA concentration was measured by reading the OD at 260 nm.

B) Oligonucleotide design

Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF, using (a) the meningococcus B sequence when available, or (b) the gonococcus/meningococcus A
25 sequence, adapted to the codon preference usage of meningococcus as necessary. Any predicted signal peptides were omitted, by deducing the 5'-end amplification primer sequence immediately downstream from the predicted leader sequence.

For most ORFs, the 5' primers included two restriction enzyme recognition sites (*Bam*HI-*Nde*I, *Bam*HI-*Nhe*I, or *Eco*RI-*Nhe*I, depending on the gene's own restriction pattern); the 3' primers included

5'-end primer tail: CGCGGATCCCATATG (*Bam*HI-*Nde*I)

CGCGGATCCGCTAGC (BamHI-NheI)

CCGGAATTCTAGCTAGC (EcoRI-NheI)

3'-end primer tail: CCCGCTCGAG (XhoI)

For ORFs 5, 15, 17, 19, 20, 22, 27, 28, 65 & 89, two different amplifications were performed to clone each ORF in the two expression systems. Two different 5' primers were used for each ORF; the same 3' *Xho*I primer was used as before:

5'-end primer tail: GGAATTCCATATGCCATGG (NdeI)

5'-end primer tail: CGGGATCC (*Bam*HI)

ORF 76 was cloned in the pTRC expression vector and expressed as an amino-terminus His-tag fusion. In this particular case, the predicted signal peptide was included in the final product. *NheI*-
15 *BamHI* restriction sites were incorporated using primers:

5'-end primer tail: GATCAGCTAGCCATATG (NheI)

3'-end primer tail: CGGGATCC (*Bam*HI)

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridized to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the whole primer, and was determined for each primer using the formulae:

$$T_E = 4 (G+C) + 2 (A+T) \quad (\text{tail excluded})$$

$$T_m = 64.9 + 0.41 (\% \text{ GC}) - 600/N \quad (\text{whole primer})$$

25 The average melting temperature of the selected oligos were 65-70°C for the whole oligo and 50-55°C for the hybridising region alone.

Table I (page 487) shows the forward and reverse primers used for each amplification. In certain cases, it will be noted that the sequence of the primer does not exactly match the sequence in the ORF. When initial amplifications were performed, the complete 5' and/or 3' sequence was not

known for some meningococcal ORFs, although the corresponding sequences had been identified in gonococcus. For amplification, the gonococcal sequences could thus be used as the basis for primer design, altered to take account of codon preference. In particular, the following codons were changed: ATA→ATT; TCG→TCT; CAG→CAA; AAG→AAA; GAG→GAA; CGA→CGC;
 5 CGG→CGC; GGG→GGC. Italicised nucleotides in Table I indicate such a change. It will be appreciated that, once the complete sequence has been identified, this approach is generally no longer necessary.

Oligos were synthesized by a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2ml NH₄OH, and deprotected by 5 hours incubation at 56°C. The oligos were precipitated by
 10 addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were then centrifuged and the pellets resuspended in either 100µl or 1ml of water. OD₂₆₀ was determined using a Perkin Elmer Lambda Bio spectrophotometer and the concentration was determined and adjusted to 2-10pmol/µl.

C) Amplification

The standard PCR protocol was as follows: 50-200ng of genomic DNA were used as a template
 15 in the presence of 20-40µM of each oligo, 400-800µM dNTPs solution, 1x PCR buffer (including 1.5mM MgCl₂), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaq, GIBCO Platinum, Pwo DNA polymerase, or Tahara Shuzo Taq polymerase).

In some cases, PCR was optimised by the addition of 10µl DMSO or 50µl 2M betaine.

After a hot start (adding the polymerase during a preliminary 3 minute incubation of the whole mix
 20 at 95°C), each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridization temperature the one of the oligos excluding the restriction enzymes tail, followed by 30 cycles performed according to the hybridization temperature of the whole length oligos. The cycles were followed by a final 10 minute extension step at 72°C.

The standard cycles were as follows:

	Denaturation	Hybridisation	Elongation
First 5 cycles	30 seconds 95°C	30 seconds 50-55°C	30-60 seconds 72°C
Last 30 cycles	30 seconds	30 seconds	30-60 seconds

	95°C	65-70°C	72°C
--	------	---------	------

The elongation time varied according to the length of the ORF to be amplified.

The amplifications were performed using either a 9600 or a 2400 Perkin Elmer GeneAmp PCR System. To check the results, 1/10 of the amplification volume was loaded onto a 1-1.5% agarose gel and the size of each amplified fragment compared with a DNA molecular weight marker.

- 5 The amplified DNA was either loaded directly on a 1% agarose gel or first precipitated with ethanol and resuspended in a suitable volume to be loaded on a 1% agarose gel. The DNA fragment corresponding to the right size band was then eluted and purified from gel, using the Qiagen Gel Extraction Kit, following the instructions of the manufacturer. The final volume of the DNA fragment was 30µl or 50µl of either water or 10mM Tris, pH 8.5.

10 D) Digestion of PCR fragments

The purified DNA corresponding to the amplified fragment was split into 2 aliquots and double-digested with:

- *NdeI/XhoI* or *NheI/XhoI* for cloning into pET-21b+ and further expression of the protein as a C-terminus His-tag fusion
- 15 – *BamHI/XhoI* or *EcoRI/XhoI* for cloning into pGEX-KG and further expression of the protein as N-terminus GST fusion.
- For ORF 76, *NheI/BamHI* for cloning into pTRC-HisA vector and further expression of the protein as N-terminus His-tag fusion.
- *EcoRI/PstI*, *EcoRI/SalI*, *SalI/PstI* for cloning into pGex-His and further expression of
- 20 the protein as N-terminus His-tag fusion

Each purified DNA fragment was incubated (37°C for 3 hours to overnight) with 20 units of each restriction enzyme (New England Biolabs) in a either 30 or 40µl final volume in the presence of the appropriate buffer. The digestion product was then purified using the QIAquick PCR purification kit, following the manufacturer's instructions, and eluted in a final volume of 30 or

- 25 50µl of either water or 10mM Tris-HCl, pH 8.5. The final DNA concentration was determined by 1% agarose gel electrophoresis in the presence of titrated molecular weight marker.

E) Digestion of the cloning vectors (pET22B, pGEX-KG, pTRC-His A, and pGex-His)

10µg plasmid was double-digested with 50 units of each restriction enzyme in 200µl reaction volume in the presence of appropriate buffer by overnight incubation at 37°C. After loading the whole digestion on a 1% agarose gel, the band corresponding to the digested vector was purified
5 from the gel using the Qiagen QIAquick Gel Extraction Kit and the DNA was eluted in 50µl of 10mM Tris-HCl, pH 8.5. The DNA concentration was evaluated by measuring OD₂₆₀ of the sample, and adjusted to 50µg/µl. 1µl of plasmid was used for each cloning procedure.

The vector pGEX-His is a modified pGEX-2T vector carrying a region encoding six histidine residues upstream to the thrombin cleavage site and containing the multiple cloning site of the
10 vector pTRC99 (Pharmacia).

F) Cloning

The fragments corresponding to each ORF, previously digested and purified, were ligated in both pET22b and pGEX-KG. In a final volume of 20µl, a molar ratio of 3:1 fragment/vector was ligated using 0.5µl of NEB T4 DNA ligase (400 units/µl), in the presence of the buffer supplied by the manufacturer.
15 The reaction was incubated at room temperature for 3 hours. In some experiments, ligation was performed using the Boehringer "Rapid Ligation Kit", following the manufacturer's instructions.

In order to introduce the recombinant plasmid in a suitable strain, 100µl *E. coli* DH5 competent cells were incubated with the ligase reaction solution for 40 minutes on ice, then at 37°C for 3 minutes, then, after adding 800µl LB broth, again at 37°C for 20 minutes. The cells were then
20 centrifuged at maximum speed in an Eppendorf microfuge and resuspended in approximately 200µl of the supernatant. The suspension was then plated on LB ampicillin (100mg/ml).

The screening of the recombinant clones was performed by growing 5 randomly-chosen colonies overnight at 37°C in either 2ml (pGEX or pTC clones) or 5ml (pET clones) LB broth + 100µg/ml ampicillin. The cells were then pelleted and the DNA extracted using the Qiagen QIAprep Spin
25 Miniprep Kit, following the manufacturer's instructions, to a final volume of 30µl. 5µl of each individual miniprep (approximately 1g) were digested with either *NdeI/XhoI* or *BamHI/XhoI* and the whole digestion loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1Kb DNA Ladder, GIBCO). The screening of the positive clones was made on the base of the correct insert size.

For the cloning of ORFs 110, 111, 113, 115, 119, 122, 125 & 130, the double-digested PCR product was ligated into double-digested vector using *EcoRI-PstI* cloning sites or, for ORFs 115 & 127, *EcoRI-SalI* or, for ORF 122, *SalI-PstI*. After cloning, the recombinant plasmids were introduced in the *E.coli* host W3110. Individual clones were grown overnight at 37°C in L-broth with 50µl/ml ampicillin.

G) Expression

Each ORF cloned into the expression vector was transformed into the strain suitable for expression of the recombinant protein product. 1µl of each construct was used to transform 30µl of *E.coli* BL21 (pGEX vector), *E.coli* TOP 10 (pTRC vector) or *E.coli* BL21-DE3 (pET vector), as described above. In the case of the pGEX-His vector, the same *E.coli* strain (W3110) was used for initial cloning and expression. Single recombinant colonies were inoculated into 2ml LB+Amp (100µg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100µg/ml) in 100ml flasks, making sure that the OD₆₀₀ ranged between 0.1 and 0.15. The flasks were incubated at 30°C into gyratory water bath shakers until OD indicated exponential growth suitable for induction of expression (0.4-0.8 OD for pET and pTRC vectors; 0.8-1 OD for pGEX and pGEX-His vectors). For the pET, pTRC and pGEX-His vectors, the protein expression was induced by addition of 1mM IPTG, whereas in the case of pGEX system the final concentration of IPTG was 0.2mM. After 3 hours incubation at 30°C, the final concentration of the sample was checked by OD. In order to check expression, 1ml of each sample was removed, centrifuged in a microfuge, the pellet resuspended in PBS, and analysed by 12% SDS-PAGE with Coomassie Blue staining. The whole sample was centrifuged at 6000g and the pellet resuspended in PBS for further use.

H) GST-fusion proteins large-scale purification.

A single colony was grown overnight at 37°C on LB+Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture in a water bath shaker and grown overnight. Bacteria were diluted 1:30 into 600ml of fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD₅₅₀ 0.8-1. Protein expression was induced with 0.2mM IPTG followed by three hours incubation. The culture was centrifuged at 8000rpm at 4°C. The supernatant was discarded and the bacterial pellet was resuspended in 7.5ml cold PBS. The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and centrifuged again. The supernatant was collected and mixed with 150µl Glutathione-Sepharose 4B resin (Pharmacia)

(previously washed with PBS) and incubated at room temperature for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml cold PBS for 10 minutes, resuspended in 1ml cold PBS, and loaded on a disposable column. The resin was washed twice with 2ml cold PBS until the flow-through reached OD₂₈₀ of 0.02-0.06. The GST-fusion protein was eluted by addition of 700µl cold Glutathione elution buffer (10mM reduced glutathione, 50mM Tris-HCl) and fractions collected until the OD₂₈₀ was 0.1. 21µl of each fraction were loaded on a 12% SDS gel using either Biorad SDS-PAGE Molecular weight standard broad range (M1) (200, 116.25, 97.4, 66.2, 45, 31, 21.5, 14.4, 6.5 kDa) or Amersham Rainbow Marker (M2) (220, 66, 46, 30, 21.5, 14.3 kDa) as standards. As the MW of GST is 26kDa, this value must be added to the MW of each GST-fusion protein.

I) His-fusion solubility analysis (ORFs 111-129)

To analyse the solubility of the His-fusion expression products, pellets of 3ml cultures were resuspended in buffer M1 [500µl PBS pH 7.2]. 25µl lysozyme (10mg/ml) was added and the bacteria were incubated for 15 min at 4°C. The pellets were sonicated for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and then separated again into pellet and supernatant by a centrifugation step. The supernatant was collected and the pellet was resuspended in buffer M2 [8M urea, 0.5M NaCl, 20mM imidazole and 0.1M NaH₂PO₄] and incubated for 3 to 4 hours at 4°C. After centrifugation, the supernatant was collected and the pellet was resuspended in buffer M3 [6M guanidinium-HCl, 0.5M NaCl, 20mM imidazole and 0.1M NaH₂PO₄] overnight at 4°C. The supernatants from all steps were analysed by SDS-PAGE.

The proteins expressed from ORFs 113, 119 and 120 were found to be soluble in PBS, whereas ORFs 111, 122, 126 and 129 need urea and ORFs 125 and 127 need guanidinium-HCl for their solubilization.

J) His-fusion large-scale purification.

A single colony was grown overnight at 37°C on a LB + Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture and incubated overnight in a water bath shaker. Bacteria were diluted 1:30 into 600ml fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD₅₅₀ 0.6-0.8. Protein expression was induced by addition of 1mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000rpm at 4°C, the supernatant was discarded and the bacterial pellet was resuspended in 7.5ml of either (i) cold

buffer A (300mM NaCl, 50mM phosphate buffer, 10mM imidazole, pH 8) for soluble proteins or (ii) buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 8.8) for insoluble proteins.

The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again.

- 5 For insoluble proteins, the supernatant was stored at -20°C, while the pellets were resuspended in 2ml buffer C (6M guanidine hydrochloride, 100mM phosphate buffer, 10mM Tris-HCl, pH 7.5) and treated in a homogenizer for 10 cycles. The product was centrifuged at 13000rpm for 40 minutes.

Supernatants were collected and mixed with 150µl Ni²⁺-resin (Pharmacia) (previously washed with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation
10 for 30 minutes. The sample was centrifuged at 700g for 5 minutes at 4°C. The resin was washed twice with 10ml buffer A or B for 10 minutes, resuspended in 1ml buffer A or B and loaded on a disposable column. The resin was washed at either (i) 4°C with 2ml cold buffer A or (ii) room temperature with 2ml buffer B, until the flow-through reached OD₂₈₀ of 0.02-0.06.

The resin was washed with either (i) 2ml cold 20mM imidazole buffer (300mM NaCl, 50mM
15 phosphate buffer, 20mM imidazole, pH 8) or (ii) buffer D (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 6.3) until the flow-through reached the O.D₂₈₀ of 0.02-0.06. The His-fusion protein was eluted by addition of 700µl of either (i) cold elution buffer A (300mM NaCl, 50mM phosphate buffer, 250mM imidazole, pH 8) or (ii) elution buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 4.5) and fractions collected until the O.D₂₈₀ was 0.1. 21µl of each
20 fraction were loaded on a 12% SDS gel.

K) His-fusion proteins renaturation

10% glycerol was added to the denatured proteins. The proteins were then diluted to 20µg/ml using dialysis buffer I (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, 2M urea, pH 8.8) and dialysed against the same buffer at 4°C for 12-
25 14 hours. The protein was further dialysed against dialysis buffer II (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was evaluated using the formula:

$$\text{Protein (mg/ml)} = (1.55 \times \text{OD}_{280}) - (0.76 \times \text{OD}_{260})$$

L) His-fusion large-scale purification (ORFs 111-129)

500ml of bacterial cultures were induced and the fusion proteins were obtained soluble in buffer M1, M2 or M3 using the procedure described above. The crude extract of the bacteria was loaded onto a Ni-NTA superflow column (Quiagen) equilibrated with buffer M1, M2 or M3 depending on the solubilization buffer of the fusion proteins. Unbound material was eluted by washing the column with the same buffer. The specific protein was eluted with the corresponding buffer containing 500mM imidazole and dialysed against the corresponding buffer without imidazole. After each run the columns were sanitized by washing with at least two column volumes of 0.5 M sodium hydroxide and reequilibrated before the next use.

10 M) Mice immunisations

20µg of each purified protein were used to immunise mice intraperitoneally. In the case of ORFs 2, 4, 15, 22, 27, 28, 37, 76, 89 and 97, Balb-C mice were immunised with Al(OH)₃ as adjuvant on days 1, 21 and 42, and immune response was monitored in samples taken on day 56. For ORFs 44, 106 and 132, CD1 mice were immunised using the same protocol. For ORFs 25 and 40, CD1 mice were immunised using Freund's adjuvant, rather than Al(OH)₃, and the same immunisation protocol was used, except that the immune response was measured on day 42, rather than 56. Similarly, for ORFs 23, 32, 38 and 79, CD1 mice were immunised with Freund's adjuvant, but the immune response was measured on day 49.

N) ELISA assay (sera analysis)

20 The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 7ml of Mueller-Hinton Broth (Difco) containing 0.25% Glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were let to grow until the OD reached the value of 0.3-0.4. The culture was centrifuged for 10 minutes at 10000rpm. The supernatant was discarded and bacteria were washed once with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 2 hours at room temperature and then overnight at 4°C with stirring. 100µl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200µl of saturation buffer (2.7% Polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed

three times with PBT. 200µl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN₃ in PBS) were added to each well and the plates incubated for 90 minutes at 37°C. Wells were washed three times with PBT. 100µl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100µl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10µl of H₂O) were added to each well and the plates were left at room temperature for 20 minutes. 100µl H₂SO₄ was added to each well and OD₄₉₀ was followed. The ELISA was considered positive when OD₄₉₀ was 2.5 times the respective pre-immune sera.

10 **O) FACScan bacteria Binding Assay procedure.**

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD₆₂₀. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA, 0.4% NaN₃) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD₆₂₀ of 0.07. 100µl bacterial cells were added to each well of a Costar 96 well plate. 100µl of diluted (1:200) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200µl/well of blocking buffer in each well. 100µl of R-Phicoerytrin conjugated F(ab)₂ goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200µl/well of blocking buffer. The supernatant was aspirated and cells resuspended in 200µl/well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan setting were: FL1 on, FL2 and FL3 off; FSC-H threshold:92; FSC PMT Voltage: E 02; SSC PMT: 474; Amp. Gains 7.1; FL-2 PMT: 539; compensation values: 0.

P) OMV preparations

Bacteria were grown overnight on 5 GC plates, harvested with a loop and resuspended in 10 ml 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes and the bacteria disrupted by sonication for 10 minutes on ice (50% duty cycle, 50% output). Unbroken cells were removed by centrifugation at 5000g for 10 minutes and the total cell envelope fraction recovered by centrifugation at 50000g at 4°C for 75 minutes. To extract cytoplasmic membrane proteins from the crude outer membranes, the whole fraction was resuspended in 2% sarkosyl (Sigma) and incubated at room temperature for 20 minutes. The suspension was centrifuged at 10000g for 10 minutes to remove aggregates, and the supernatant further ultracentrifuged at 50000g for 75 minutes to pellet the outer membranes. The outer membranes were resuspended in 10mM Tris-HCl, pH8 and the protein concentration measured by the Bio-Rad Protein assay, using BSA as a standard.

Q) Whole Extracts preparation

Bacteria were grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

15 R) Western blotting

Purified proteins (500ng/lane), outer membrane vesicles (5µg) and total cell extracts (25µg) derived from MenB strain 2996 were loaded on 15% SDS-PAGE and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, in transferring buffer (0.3 % Tris base, 1.44 % glycine, 20% methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

S) Bactericidal assay

MC58 strain was grown overnight at 37°C on chocolate agar plates. 5-7 colonies were collected and used to inoculate 7ml Mueller-Hinton broth. The suspension was incubated at 37°C on a nutator and let to grow until OD₆₂₀ was 0.5-0.8. The culture was aliquoted into sterile 1.5ml Eppendorf

tubes and centrifuged for 20 minutes at maximum speed in a microfuge. The pellet was washed once in Gey's buffer (Gibco) and resuspended in the same buffer to an OD₆₂₀ of 0.5, diluted 1:20000 in Gey's buffer and stored at 25°C.

50µl of Gey's buffer/1% BSA was added to each well of a 96-well tissue culture plate. 25µl of diluted mice sera (1:100 in Gey's buffer/0.2% BSA) were added to each well and the plate incubated at 4°C. 25µl of the previously described bacterial suspension were added to each well. 25µl of either heat-inactivated (56°C waterbath for 30 minutes) or normal baby rabbit complement were added to each well. Immediately after the addition of the baby rabbit complement, 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 0). The 96-well plate was incubated for 1 hour at 37°C with rotation and then 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 1). After overnight incubation the colonies corresponding to time 0 and time 1 hour were counted.

Table II (page 493) gives a summary of the cloning, expression and purification results.

Example 1

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 1>:

```

1  ATGAAACAGA CAGTCAA.AT GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51  GAACCGACCG GTGTGGNCGG ATGACGTATC GGATTTTCGG GAAACTTGC
101 A.GCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
151 TAT.TACAAA GGCAGCGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
201 GTATCGGCAG CCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251 GCTGGATGTA TGCCAACGGG CGCGC.GTGC GCCAAGATGA TACCGAAGCG
301 GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351 CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
401 TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
251 451 GCCCAAAACA ATTTGGGCGT GATGTATGCC GAAAGANCGC GCGTGCGCCA
501 AGACCG...
```

This corresponds to the amino acid sequence <SEQ ID 2; ORF37>:

```

1  MKQTVXMLAA ALIALGLNRP VWXDDVSDFR ENLXAAAQGN AAAQYNLGAM
51  YXQTRVRRD DAEAVRWYRQ PAEQGLAQAO YNLGWMYANG RXVRQDDTEA
101 VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAG
151 AQNNLGVMYA ERXRVQRD...
```

Further work revealed the complete nucleotide sequence <SEQ ID 3>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
51  GAACCGAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAACTTGC
101 AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
151 TATTACAAAG GACGCGGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
201 GTATCGGCAG CCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251 GCTGGATGTA TGCCAACGGG CGCGGCGTGC GCCAAGATGA TACCGAAGCG
301 GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351 CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
401 TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
451 GCCCAAAACA ATTTGGGCGT GATGTATGCC GAAAGACGCG GCGTGCGCCA
501 AGACCGCGCC CTTGCACAAG AATGGTTTGG CAAGGCTTGT CAAAACGGAG
551 ACCAAGACGG CTGCGACAAT GACCAACGCC TGAAGGCGGG TTATTGA
```

This corresponds to the amino acid sequence <SEQ ID 4; ORF37-1>:

```

1  MKQTVKWLAA ALIALGLNRA VWADDVSDFR ENLQAAQGN AAAQYNLGAM
51 YYKGRGVRD DAEAVRWYRQ AAEQGLAQAO YNLGWMYANG RGVRRDDTEA
101 VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAAQ
151 AQNNLGVMYA ERRGVRQDRA LAQEWFGKAC QNGDQDGCND DQRLKAGY*

```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 5>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CTTGGGCTT
51 GAACCAAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAACTTGC
101 AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AAAACAATTT GGGCGTGATG
151 TATGCCGAAA GACGCGGCGT GCGCCAAGAC CGCGCCCTTG CACAAGAATG
201 GCTTGGCAAG GCTTGTCAAA ACGGATACCA AGACAGCTGC GACAATGACC
251 AACGCCTGAA AGCGGTTTAT TGA

```

This encodes a protein having amino acid sequence <SEQ ID 6; ORF37a>:

```

1  MKQTVKWLAA ALIALGLNQA VWADDVSDFR ENLQAAQGN AAQNNLGVM
151 YAERRGVRQD RALAQEWLGK ACQNGYQDSC DNDQRLKAGY *

```

The originally-identified partial strain B sequence (ORF37) shows 68.0% identity over a 75aa overlap with ORF37a:

```

20      orf37.pep      10      20      30      40      50      60
      MKQTVXMLAAALIALGLNRPVWXXDDVSDFR ENLXAAAQGNAAAQYNLGAMYXQRTVRD
      orf37a          10      20      30      40      50      60
      MKQTVKWLAAALIALGLNQA VVWADDVSDFR ENLQAAQGNAAAQNNLGVMYAERRGVRQD
25      orf37.pep      70      80      90      100     110     120
      DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG
      orf37a          70      80      90
      RALAQEWLGKACQNGYQDSCDNDQRLKAGYX

```

30 Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 7>:

```

1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CTTGGGCTT
51 GAACCAAGCG GTGTGGGCGG GTGACGTATC GGATTTTCGG GAAACTTGC
101 AGGCGGCaGA ACAGGGAAAT GCAGCAGCCC AATTCAATTT GGGCGTGATG
151 TATGAAAATG GACAAGGAGT TCGTCAAGAT TATGTACAGG CAGTGCAGTG
35 201 GTATCGCAAG GCTTCAGAAC AAGGGGATGC CCAAGCCCAA TACAATTGG
251 GCTTGATGTA TTACGATGGA CGCGGCGTGC GCCAAGACCT TGCCTCGCT
301 CAACAATGGC TTGGCAAGGC TTGTCAAAC GGAGACCAA ACAGCTGCCA
351 CAATGACCAA CGCTGAAGG CGGGTTATTA A

```

This encodes a protein having amino acid sequence <SEQ ID 8; ORF37ng>:

```

40 1  MKQTVKWLAA ALIALGLNQA VWAGDVSDFR ENLQAAEQGN AAQFNLGVM
51 YENGQGVQRD YVQAVQWYRK ASEQGDAAQ YNLGLMYDGR RGVRRDLALA
101 QWLKGKACQN GDQNSCDNDQ RLKAGY*

```

The originally-identified partial strain B sequence (ORF37) shows 64.9% identity over a 111aa overlap with ORF37ng:

```

45      orf37.pep      MKQTVXMLAAALIALGLNRPVWXXDDVSDFR ENLXAAAQGNAAAQYNLGAMYXQRTVRD      60
      orf37ng          MKQTVKWLAAALIALGLNQA VVWAGDVSDFR ENLQAAEQGNAAAQFNLGVMYENGQGVQRD      60
50      orf37.pep      DAEAVRWYRQPAEQGLAQAYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG      120
      orf37ng          YVQAVQWYRKASEQGDAAQAYNLGLMYDGRGVRRDLALAQQWLKGKACQNGDQNSCDNDQ      120
      orf37.pep      VIYAEGRGVRQDDVEAVRWFRQAAAQGVAAQNNLGVMYAERXVRQD      168
55      orf37ng          RLKAGY                                          126

```

The complete strain B sequence (ORF37-1) and ORF37ng show 51.5% identity in 198 aa overlap:

		10	20	30	40	50	60
	orf37-1.pep	MKQTVKWLAAALIALGLNRAVWADDVSD	FREN	LQAA	QGNAA	QYNLG	AMYYKGRGVR
5	orf37ng	MKQTVKWLAAALIALGLNRAVWAGDVSD	FREN	LQAA	QGNAA	QYNLG	VMYENGQGV
		10	20	30	40	50	60
	orf37-1.pep	DAEAVRWYRQAAEQGLAQYINLGW	MYANG	RGRQDD	TEAVR	WYRQAA	QGVVQAQYN
10	orf37ng	YVQAVQWYRKASEQGDAAQYINL	GLMYD	GRGVRQD	-----	-----	-----
		70	80	90			
15	orf37-1.pep	VIYAEGRGVRQDDVEAVRWFRQAA	QGVQA	QNNLG	VMYA	ERRGV	RQDRALAQEW
	orf37ng	-----	-----	-----	-----	-----	-----
		130	140	150	160	170	180
20	orf37-1.pep	QNGDQDGCNDQRLKAGYX					
	orf37ng	QNGDQNSCDNDQRLKAGYX					
		110	120				

- 25 Computer analysis of these amino acid sequences indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF37-1 (11kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 1A shows the results of affinity purification of the GST-fusion protein, and Figure 1B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), FACS analysis (Figure 1C), and a bactericidal assay (Figure 1D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is a useful immunogen.

- 35 Figure 1E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF37-1.

Example 2

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 9>:

40 TTCGGCGA CATCGGCGGT TTGAAGGTCA ATGCCCCCGT CAAATCCGCA
GGCGTATTGG TCGGGCGCGT CGGCGCTATC GGAATTGACC CGAAATCCTA
TCAGGCGAGG GTGCGCCTCG ATTTGGACGG CAAGTATCAG TTCAGCAGCG
ACGTTTCCGC GCAAATCCTG ACTTCSGGAC TTTTGGGCGA GCAGTACATC
GGGCTGCAGC AGGGCGGCGA CACGGAAC CTTGCTGCCG GCGACACCAT
CTCCGTAACC AGTTCTGCAA TGGTTCTGGA AAACCTTATC GGCAAATTC
45 TGACGAGTTT TGCCGAGAAA AATGCCGACG GCGGCAATGC GGAAAAAGCC
GCCGAATAA

This corresponds to the amino acid sequence <SEQ ID 10>:

1 FGDIGGLKVN APVKSAGVLV GRVGAIGLDP KSYQARVRLD LDGKYQFSSD
51 VSAQILTSLG LGEQYIGLQQ GGDENLAAG DTISVTSSAM VLENLIGKFM

101 TSFAEKNADG GNAEKAEE*

Computer analysis of this amino acid sequence gave the following results:

Homology with a hypothetical *H.influenzae* protein (ybrd.haein; accession number p45029)

SEQ ID 9 and ybrd.haein show 48.4% aa identity in 122 aa overlap:

```

5      20      30      40      50      60      70
ybrd.h LGIGALVFLGLRVANVQGFATKSYTVTATFDNIGGLKVRAPLKIGGVVIGRVSAITLDE
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
N.m      FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
              10      20      30

10     80      90      100     110     120     130
ybrd.h KSYLPKVSIAINQYNEIPESSLSIKTSGLLGEQYIALTMGFDDGDTAMLKNGSQIQDT
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
N.m      KSYQARVRLDLGKY-QFSSDVSAQILTSGLLGEQYIGLQGG---GDTENLAAGDTISVT
              40      50      60      70      80

15     140     150     160
ybrd.h TSAMVLEDLIGQFL--YGSKKSDGNEKSESTEQ
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
N.m      SSAMVLENLIGKFMTSFAEKNADGGNAEKAEX
              90      100     110     120

```

Homology with a predicted ORF from *N.gonorrhoeae*

SEQ ID 9 shows 99.2% identity over a 118aa overlap with a predicted ORF from *N. gonorrhoeae*:

```

25     20      30      40      50      60      70
ybrd      GAAAVAFLAFRVAGGAFFGGSDKTYAVYADFGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
N.m      FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
              10      20      30

30     80      90      100     110     120     130
ybrd      KSYQARVRLDLGKYQFSSDVSAQILTSGLLGEQYIGLQGGDTENLAAGDTISVTSSAM
N.m      KSYQARVRLDLGKYQFSSDVSAQILTSGLLGEQYIGLQGGDTENLAAGDTISVTSSAM
              40      50      60      70      80      90

35     140     150     160
ybrd      VLENLIGKFMTSFAEKNADGGNAEKAEX
N.m      VLENLIGKFMTSFAEKNADGGNAEKAEX
              100     110     120

```

The complete ybrd *H.influenzae* sequence has a leader sequence and it is expected that the full-length homologous *N.meningitidis* protein will also have one. This suggests that it is either a membrane protein, a secreted protein, or a surface protein and that the protein, or one of its epitopes, could be a useful antigen for vaccines or diagnostics.

Example 3

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 11>:

```

50      1  ..ATTTGATAT ACCTCATCCG CAAGAATCTA GGTTCGCCCG TCTTCTTCTT
      51  TCAGGAACGC CCCGGAAGG ACGGAAAACC TTTTAAATG GTCAAATCC
      101  GTTCCATCGC CGACGGCTTG TATTCAGACG GCATTCCGCT GCCCGACGGA
      151  GAACGCCTGA CACCGTTCGG CAAAAAACTG CGTGCCGcCA GTWTGGACGA
      201  ACTGCCTGAA TTATGGAATA TCTTAAAGG CGAGATGAGC CTGGTCGGCC
      251  CCCGCCCGCT GCTGATGCAA TATCTGCCGC TGTACGACAA CTTCCAAAC
      301  CGCCGCCACG AATGAAACC CGGCATTACC GGCTGGGCGC AGGTCAACGG

```

5
10
351 GCGCAACGCG CTTTCGTGGG ACGAAAAATT CGCCTGCGAT GTTTGGTATA
401 TCGACCACTT CAGCCTGTGC CTCGACATCA AAATCCTACT GCTGACGGTT
451 AAAAAAGTAT TAATCAAGGA AGGGATTTCC GCACAGGGCG AACAA .aCCAT
501 GCCCCCTTTC ACAGGAAAAAC GCAAACTCGC CGTCGTCCGT GCGGGCGGAC
551 ACGGAAAAGT CGTTGCCGAC CTGCGCCCG CACTCGGCCG GTACAGGGAA
601 ATCGTTTTTC TGGACGACCG CGCACAAGGC AGCGTCAACG GCTTTTCCGT
651 CATCGGCACG ACGCTGCTGC TTGAAAAACAG TTTATCGCCC GAACAATACG
701 ACGTCGCCGT CGCCGTCCGC AACAAACCGCA TCCGCGGCCA AATCGCCGAA
751 AAAGCCGCCG CGCTCGGCTT CGCCTGCCC GTACTGGTTC ATCCGACGCG
801 GACCGTCTCG CTTTCTGCAA CAGTCGGACA AGGCAGCGTC GTTATGGCGA
851 AAGCGGTCG..

This corresponds to the amino acid sequence <SEQ ID 12; ORF3>:

15
1 .. ILIYLIRKLNL GSPVFFQER PGKDGKPFKM VKFRSMRDGL YSDGILPDPG
51 ERLTPFGKKL RAASXDELPE LWNILKGEMS LVGPRPLLMQ YLPLYDNFQN
101 RRHEMKPGIT GWAQVNGRNA LSWDEKFACD VWYIDHFSLC LDIKILLTLV
151 KKVLIKEGIS AQGEXTMPFF TGRKRLAVVG AGGHGKVVD LAAALGRYRE
201 IVFLDDRAQG SVNGFSVIGT TLLLENSLSP EQYDVAVAVG NNRIRRQIAE
251 KAAALGFALP VLVHPDATVS PSATVGQGSV VMAKAV..

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 13>:

20
1 ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG
51 ACTGATTTTC CTCTCGCCAG TATTTTGTAT TTTGATATAC CTCATCCGCA
101 AGAATCTAGG TTCGCCGCTC TTCTTCTTTC AGGAACGCCC CGGAAAGGAC
151 GGAAAACTTT TTTAAATGGT CAAATTCCGT TCCATGCGCG ACGCGCTTGA
201 TTCAGACGGC ATTCCGCTGC CCGACGGAGA ACGCCTGACA CCGTTCGGCA
251 AAAAAGGCG TGCCGCCAGT TTGGACGAAC TGCTGAATT ATGGAATATC
301 TTTAAAGGCG AGATGAGCCT GGTCCGCCCG CGCCCGCTGC TGATGCAATA
351 TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAACCCG
401 GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC
451 GAAAAATTCG CCTGCGATGT TTGGTATATC GACCACTTCA GCCTGTGCCT
30 501 CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAGGAAG
551 GGATTTCCGC ACAGGGCGAA GCCACCATGC CCCCTTTCAC AGGAAAACGC
601 AAATCGCCG TCGTCGGTGC GGGCGGACAC GGAAGTTCG TTGCCGACCT
651 TGCCGCCGCA CTCGGCCGGT ACAGGGAAT CGTTTTTCTG GACGACCGCG
701 CACAAGGCAG CGTCAACGGC TTTTCCGTCA TCGGCACGAC GCTGCTGCTT
35 751 GAAACAGTT TATCGCCGA ACAATACGAC GTCGCCGTCG CCGTCGGCAA
801 CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCGCCGCG CTCGGCTTCG
851 CCCTGCCCGT TCTGGTTCAT CCGGACGCGA CCGTCTCGCC TTCTGCAACA
901 GTCGGACAAG GCAGCGTCGT TATGGCGAAA GCCGTCGTAC AGGCAGGCAG
951 CGTATTGAAA GACGGCGTGA TTGTGAACAC TGCCGCCACC GTCGATCAGC
40 1001 ACTGCCTGCT TAACGCTTTC GTCCACATCA GCCCAGGCGC GCACCTGTGC
1051 GGCAACACGC ATATCGGCGA AGAAAGCTGG ATAGGCACGG GCGCGTCGAG
1101 CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGGCGCAG
1151 TCGTCGTACG CGACGTTTCA GACGGCATGA CCGTCGCGGG CAATCCGGCA
1201 AAGCCGCTGC CGCGCAAAAA CCCCAGAGCC TCGACAGCAT AA

45 This corresponds to the amino acid sequence <SEQ ID 14; ORF3-1>:

50
1 MSKFFKRLFD IVASASGLIF LSPVFLILY LIRKNLGSFV FFFQERPGKD
51 GKPFKMKVFR SMRDALSDG IPLPDGERLT PFGKKLRAAS LDELPELWNI
101 LKGEMSLVGP RPLLMQYLPL YDNFQNRHE MKPGITGWAQ VNGRNALSWD
151 EKFCADVWYI DHFSLCLDIK ILLLTVKKVL IKEGISAQGE ATMPPTGKR
201 KLAVVGAGGH GKVVADLAAA LGRYREIVFL DDRAQGSVNG FSVIGTLLL
251 ENSLSPEQYD VAVAVGNRRI RRQIAEKAAA LGFALPVLVH PDATVSPSAT
301 VGQGSVMAK AVVQAGSVLK DGVIINTAAT VDHDCLLNAF VHISPGAHL
351 GNTHIGESW IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA
401 KPLPRKNPET STA*

55 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF3 shows 93.0% identity over a 286aa overlap with an ORF (ORF3a) from strain A of *N. meningitidis*:

10

20

30

	1	ATGAGTAAAT	TCTTCAAACG	CCTGTTTGAC	ATTGTTGCCT	CCGCTCGGG
	51	ACTGATTTTC	CTCTGCCAG	TATTTTGTAT	TTTGATATAC	CTCATCCGCA
	101	AGAATCTGGG	TTGCGCCGTC	TTCTTCTTTC	AGGAACGCC	CGGAAAGGAC
	151	GGAAACCTT	TTAAATGGT	CAAATTCGT	TCCATGCACG	ACGCGCTTGA
40	201	TTGAGACGGC	ATTCTGCTGC	CCGACGGASA	ACGCCTGACA	CCGTTCGGCA
	251	AAAACTGCG	TGCGCCAGT	TTGGACGAAC	TGCCCCGAAT	GTGGAACGTC
	301	CTCAAAGGCG	ACATGAGCCT	GSTCGGCCCC	CGCCCGCTGC	TGATGCAATA
	351	TCTGCGCGTG	TACGACAACT	TCCAAAACCG	CCGCCACGAA	ATGAAACCGG
	401	GCATTACCGG	CTGGGCGCAG	GTCAACGGGC	GCAACGCGCT	TTCTGTTGGAC
45	451	GAACGCTTCG	CATGCGACAT	CTGGTATATC	GACCACTTCA	GCCTGTGCCT
	501	CGACATCAAA	ATCCTACTGC	TGACGGTTAA	AAAAGTATTA	ATCAAAGAAG
	551	GGATTTCGCG	ACAGGGCGAA	GCCACCATGC	CCCTTTTCAC	AGGAAAACCG
	601	AAACTTGGCG	TCGTGCGTGC	GGGCGGACAC	GGCAAAGTCG	TTGCCGAGCT
	651	TGCCGCGCGA	CTCGGCACAT	ACGCGCAAAT	CGTTTCTTCTG	GACGACCCGG
50	701	TCCAAGGCGA	CGTCAACGGC	TTCCCCGTCA	TCGGCACGAC	GCTGCTGCTT
	751	GAAAACAGTT	TATCGCCCGA	ACAATTCGAC	ATCGCGCTCG	CCGTTCGGCAA
	801	CAACCGCATC	CCTCGCCAAA	TCGCCGAAAA	AGCCGCGCGG	CTCGGCTTCG
	851	CCCTGCCCGT	CGTGATTCAT	CCGGACTCGA	CCGTCTCGCC	TTCTGCAACA
	901	GTCGGACAAG	GCGGCGTCGT	TATGGCGAAA	GCCGTCGTAT	AGGCTGACAG
55	951	CGTATTGAAA	GACGGCGTAA	TTGTGAACAC	TGCCGCCACC	GTCGATCACG
	1001	ATTGCCCTGCT	TGATGCTTTC	GTCCACATCA	GCCCCGGCGC	GCACCTGTCTG
	1051	GGCAACACGC	GTATCGGCGA	AGAAAGCTGG	ATAGGCACAG	GCGCGTGCAG
	1101	CCGCCAGCAG	ATCCGTATCG	GCAGCCGCGC	AACCATTTGA	GCGGGCCGAG
	1151	TCGTGCTGCG	GCGACTTTCA	GACGGCATGA	CCGTTCGCGG	CACCCGCGCA
60	1201	AAACCATTGG	CAGGCAAAAA	TACCGAGACC	CTGCGGTCGT	AA

```

1 MSKFFKRLFD IVASAGSLIF LSPVFLILIY LIRKNLGSPV FFFQERPGKD
51 GKPFKMVKFR SMHDAALSDG ILLPDGERLT PFGKKLRAAS LDELPELWNV
101 LKGDSL VGPR RPLLMQYLPL YDNFQNRHE MKPGITGWAQ VNGRNLASWD
151 ERFACDIWYI DHFSLCLDIK ILLLTVKKVL IKGISAQGE ATPMPTGKR
201 KLAUVGAGGH GKVVAELAAA LGTYGEIVFL DDERVQGSVNG FPIVGTLLLL
251 ENSLSPEQFD IYAVAVGNMRI RRQIAEKAAA LGFALPVLH PDSTVSPSAT

```


301 VGQGGVVMK AVVQADSVLK DGVIVNTAAT VDHDCLLDAF VHISPGAHL S
 351 GNTRIGESW IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA
 401 KPLAGKNTET LRS*

Two transmembrane domains are underlined.

5 ORF3-1 shows 94.6% identity in 410 aa overlap with ORF3a:

		10	20	30	40	50	60
	orf3a.pep	MSKFFKRLFDIVASASGLIFLSPVFLILYLRKKNLGS	PVFFQERPGKDGKPFKMKVFR				
	orf3-1	MSKFFKRLFDIVASASGLIFLSPVFLILYLRKKNLGS	PVFFQERPGKDGKPFKMKVFR				
10		10	20	30	40	50	60
	orf3a.pep	SMHDALDSGILLPDGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGPRPLLMQYLPL					
	orf3-1	SMRDALDSGILPLPDGERLTPFGKKLRAASLDELPELWNILKGEMSLVGPRPLLMQYLPL					
15		70	80	90	100	110	120
	orf3a.pep	YDNFQNRHMKPGITGWAQVNGRNALS	WDERFACDIWYIDHFSCLDIKILLTVKKVL				
	orf3-1	YDNFQNRHMKPGITGWAQVNGRNALS	WDEKFACDVWYIDHFSCLDIKILLTVKKVL				
20		130	140	150	160	170	180
	orf3a.pep	IKEGISAQGEATMPFFT	GKRKLAVVGAGGHGKVVAAALGTYGEIVFLDDRVQGSVNG				
	orf3-1	IKEGISAQGEATMPFFT	GKRKLAVVGAGGHGKVADLAAALGRYREIVFLDDRAQGSVNG				
25		190	200	210	220	230	240
	orf3a.pep	FPVIGTTLLENLSPEQFDIAVAVGNNRIRRQIAEKAAALGFALPVLHPDSTVSPSAT					
	orf3-1	FSVIGTTLLENLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT					
30		250	260	270	280	290	300
	orf3a.pep	VGQGGVVMKAVVQADSVLKDGVIVNTAATVDHDCLLDAFVHISPGAHL	SGNTRIGESW				
	orf3-1	VGQGSVVMKAVVQAGSVLKDGVIVNTAATVDHDCLLNAFVHISPGAHL	SGNTHIGESW				
35		310	320	330	340	350	360
	orf3a.pep	IGTACSRQQIRIGSRATIGAGAVVVRDVS	DGMTVAGNPAKPLAGKNTETLRX				
	orf3-1	IGTACSRQQIRIGSRATIGAGAVVVRDVS	DGMTVAGNPAKPLPRKNPETSTAX				
40		370	380	390	400	410	
	orf3a.pep	IGTACSRQQIRIGSRATIGAGAVVVRDVS	DGMTVAGNPAKPLPRKNPETSTAX				
	orf3-1	IGTACSRQQIRIGSRATIGAGAVVVRDVS	DGMTVAGNPAKPLPRKNPETSTAX				

Homology with hypothetical protein encoded by yvfc gene (accession Z71928) of *B. subtilis*

ORF3 and YVFC proteins show 55% aa identity in 170 aa overlap (BLASTp):

50	ORF3	3	IYLIRKNLGS	PVFFQERPGKDGKPFKMKVFR	SMRDGLYSDGIPLPDGERLTPFGKKLRA	62
	yvfc	27	IAVVR	LKIGSPVFFQVRPGLHGKPF	TLTKFRTMTDERDSKGNLLPDEVRLTKTGRLIRK	86
55	ORF3	63	ASXDELPELWNILKGEMSLVGPRPLLMQYLPLYDNFQNRHMKPGITGWAQVNGRNALS	122		
	yvfc	87	LSIDELPQLNVLKGDLSLVGPRPLMDYLPLYTEQARRHEVKPGITGWAQINGRNAIS	146		
60	ORF3	123	WDEKFACDVWYIDHFSCLDXXXXXXXXXXXXXXXXXEGISAQEXTMPFFT	172		
	yvfc	147	WEKKFELDVWYVDNWSFFLDLKLCLTVRKVLVSEGIQQTNHVTAERFTG	196		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF3 shows 86.3% identity over a 286aa overlap with a predicted ORF (ORF3.ng) from *N. gonorrhoeae*:

5	orf3	ILIIYLIRKKNLGSPVFFFQERPGKDGKPFKMKVFR	34
	orf3ng	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLIRKKNKGSVFFIRERPGKDGKPFKMKVFR	60
10	orf3	SMRDGLYSDGIPDPGERLTDFGKKLRAASXDELPELWNILKGEMSLVGPRPLLMQYLPL	94
	orf3ng	SMRDALDSGIPDPSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLLMQYLPL	120
15	orf3	YDNFQNRHRHEMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSCLCDIKILLTVKKVL	154
	orf3ng	YNKFQNRHRHEMKPGITGWAQVNGRNALSWDEKFSVDVWYTDNFSFWLDMKILFLTVKKVL	180
20	orf3	IKEGISAQGEEXTMPPTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG	214
	orf3ng	IKEGISAQGEATMPPFAGNRKLAVIGAGGHGKVVAEALAAALGTGYEIVFLDDRTQGSVNG	240
25	orf3	FSVIGTLLLENSLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT	274
	orf3ng	FPVIGTLLLENSLSPEQFDITVAVGNNRIRRQITENAAALGFALPVLVHPDATVSPSAI	300
30	orf3	VGQGSVVMKAV	286
	orf3ng	IGQGSVVMKAVVQAGSVLKDGIVNTAATVDHDCLLDAFVHISPGAHLNTRIGEEESR	360

The complete length ORF3ng nucleotide sequence <SEQ ID 17> is:

1	ATGAGTAAAG	CCGTCAAACG	CCTGTTTCGAC	ATCATCGCAT	CCGCATCGGG
51	GCTGATGTGC	CTGTGCCCGG	TGTTTTTGGT	TTTAATATAC	CTCATCCGCA
101	AAACTTAGG	TTCCGCCGTC	TTCTTCattc	GGGAACGCCc	cgGAAAGGAc
151	ggaaaacCTT	TTAAATGGT	CAAATTCCTG	TCCAtgcgcg	acgcgcttGA
201	TTGAGACGGC	ATTCCGCTGC	CCGATAGCGA	ACGCCTGACC	GATTTCGGCA
251	AAAAATTACG	CGCCACCACT	TTGGACGAAC	TTCTCTGAAT	ATGGAATGTC
301	CTCAAAGGCG	AGATGAGCCT	GGTCGGCCCC	CGCCCGCTTT	TGATGCAGTA
351	TCTGCCGCTT	TACAACAAAT	TTCAAACACG	CCGCCACGAA	ATGAAACCCG
401	GCATTACCGG	CTGGGCGCAG	GTCAACGGGC	GCAACGCGCT	TTCGTGGGAC
451	GAAAAGTTCT	CCTGCGATGT	TTGGTACACC	GACATTTTCA	GCTTTTGGCT
501	GGATATGAAA	ATCCTGTTTC	TGACAGTCAA	AAAAGTCTTG	ATTAAAGAAG
551	GCATTTCGGC	GCAAGGGGAA	GCCACCATGC	CCCCTTTCGC	GGGGAATCGC
601	AAACTCGCCG	TTATCGGCGC	GGGCGGACAC	GGCAAAGTCG	TTGCCGAGCT
651	TGCCGCCGCA	CTCGGCACAT	ACGGCGAAAT	CGTTTTTCTG	GACGACCGCA
701	CCCAAGGCAG	CGTCAACGGC	TTCCCGCTCA	TCGGCACGAC	GCTGTGCTT
751	GAAAACAGTT	TATCGCCGGA	ACAATTCGAC	ATCACCGTCG	CCGTCCGCAA
801	CAACCGCATC	CGCCGCCAAA	TCACCGAAAA	CGCCGCCGCG	CTCGGCTTCA
851	AACTGCCCGT	TCTGATTCAT	CCCGACGCGA	CCGTCTCGCC	TTCTGCAATA
901	ATCGGACAAG	GCAGCGTCGT	AATGGCGAAA	GCCGTCTGAC	AGGCCGGCAG
951	CGTATTGAAA	GACGGCGTGA	TTGTGAACAC	TGCCGCCACC	GTCGATCACG
1001	ACTGCCTGCT	TGACGCTTTC	GtccaCATCA	GCCCGGGCGC	GCACCTGTCTG
1051	GGCAACACGC	GTATCGGCGA	AGAAAGCCGG	ATAGGCACGG	GCGCGTGCAG
1101	CCGCCAGCAG	ACAACCGTCG	GCAGCGGGGT	TACCgcgGT	GCAGGGcgCG
1151	TTATCGTATG	CGACATCCCG	GACGGCATGA	CCGTCCGGGG	CAACCCGGCA
1201	AAGCCCTTA	CGGGCAAAAA	CCCCAAGACC	GGGACGGCAT	AA

This encodes a protein having amino acid sequence <SEQ ID 18>:

55	1	MSKAVKRLFD	IIASASGLIV	LSPVFLVLIY	LIRKKNLGSPV	FFIRERPGKD
	51	GKPFKMKVFR	SMRDALDSG	IPLPDSERLT	DFGKKLRATS	LDELPELWNV
	101	LKGEMSLVGP	RPLLMQYLPL	YNKFQNRHE	MKPGITGWAQ	VNGRNALSWD
	151	EKFSCDVWYT	DNFSFWLDMK	ILFLTVKKVL	IKEGISAQGE	ATMPPFAGNR
	201	KLAVIGAGGH	GKVVAELAAA	LGTYGEIVFL	DDRTQGSVNG	FPVIGTLLLL
	251	ENSLSPQFD	ITVAVGNRI	RRQITENAAA	LGFKLPVLIH	PDATVSPSAI
60	301	IGQGSVVMK	AVVQAGSVLK	DGVIVNTAAT	VDHDCLLDAF	VHISPGAHLN
	351	GNTRIGEEESR	IGTGACSRQQ	TTVSGSVTAG	AGAVIVCDIP	DGMTVAGNPA
	401	KPLTGKNPKT	GTA*			

This protein shows 86.9% identity in 413 aa overlap with ORF3-1:

		10	20	30	40	50	60
	orf3-1.pep	MSKFFKRLFDIVASASGLIFLSPVFLIYLIYLRKNLGS	PVFFQERPGKDGKPFKMKVFR				
5	orf3ng	MSKAVKRLFDIIASASGLIVLSPVFLVLIYLRKNLGS	PVFFIRERPGKDGKPFKMKVFR				
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf3-1.pep	SMRDALDSGIPDPGERLTPFGKKLRAASLDELPELWNILKGEMSLVGPRPLLMQYLPL					
10	orf3ng	SMRDALDSGIPDPGERLTPFGKKLRAASLDELPELWNILKGEMSLVGPRPLLMQYLPL					
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf3-1.pep	YDNFQNRHHEMKPGITGWAQVNGRNASWDEKFCADVWYIDHFSCLDIKILLTVKKVL					
15	orf3ng	YDKFQNRHHEMKPGITGWAQVNGRNASWDEKFCADVWYIDHFSCLDIKILLTVKKVL					
		130	140	150	160	170	180
		190	200	210	220	230	240
	orf3-1.pep	IKEGISAQGEATMPFPGKRLAVVAGGHHGKVVADLAAALGRYREIVFLDDRAQGSVNG					
20	orf3ng	IKEGISAQGEATMPFPGKRLAVVAGGHHGKVVADLAAALGRYREIVFLDDRAQGSVNG					
		190	200	210	220	230	240
		250	260	270	280	290	300
	orf3-1.pep	FSVIGTTLLENLSLSPQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT					
25	orf3ng	FPVIGTTLLENLSLSPQYDVAVAVGNNRIRRQITENAAALGFALPVLVHPDATVSPSAT					
		250	260	270	280	290	300
		310	320	330	340	350	360
	orf3-1.pep	VGQGSVVMKAVVQAGSVLKDGVIVNTAATVDHDCLLNAFVHISPGAHLNTHIGEEESW					
30	orf3ng	IGQGSVVMKAVVQAGSVLKDGVIVNTAATVDHDCLLNAFVHISPGAHLNTHIGEEESW					
		310	320	330	340	350	360
		370	380	390	400	410	
	orf3-1.pep	IGTGACSRQQIRIGSRATIGAGAVVVRDVS DGMTVAGNPAKPLPRKNPETSTAX					
35	orf3ng	IGTGACSRQQITVGSVTAGAGAVIVCDIPDGMTVAGNPAKPLTGKNPKTGTAX					
		370	380	390	400	410	

In addition, ORF3ng shows significant homology with a hypothetical protein from *B. subtilis*:

45	gnl PID e238668 (271928) hypothetical protein [Bacillus subtilis]
	>gi 1945702 gnl PID e313004 (294043) hypothetical protein [Bacillus subtilis]
	>gi 2635938 gnl PID e1186113 (299121) similar to capsular polysaccharide biosynthesis [Bacillus subtilis] Length = 202
	Score = 235 bits (594), Expect = 3e-61
	Identities = 114/195 (58%), Positives = 142/195 (72%)
50	Query: 5 VKRLFDIIASASGLIVLSPVFLVLIYLRKNLGS
	+KRLFD+ A+ L S + L I ++R +GSPVFF + RFG GKPF + KFR+M D
	Sbjct: 3 LKRLFDLTAAIFLLCCTSVIILFTIAVRLKIGSPVFFKQVRPGLHGKPFITLYKFRMTD 62
55	Query: 65 ALDSGIPDPGERLTPFGKKLRAASLDELPELWNILKGEMSLVGPRPLLMQYLPLYNKF 124
	DS G LPD RLT G+ +R S+DELP+L NVLKG++SLVGPRPLLM YLPLY +
	Sbjct: 63 ERDSKGNLLPDEVRLTKTGRLIRKLSIDELPQLLNVLKGDLSLVGPRPLLMQYLPLYTEK 122
60	Query: 125 QNRHHEMKPGITGWAQVNGRNASWDEKFCADVWYIDHFSCLDIKILLTVKKVLKEG 184
	Q RRHE+KPGITGWAQ+NGRNA+SW++KF DVWY DN+SF+LD+KIL LTV+KVL+ EG
	Sbjct: 123 QARRHEVMPGITGWAQINGRNAISWEKKFELDVWYVDNWSFLLDLKILCLTVRKVLVSEG 182
	Query: 185 ISAQGEATMPFAGN 199
	I T F G+
65	Sbjct: 183 IQQTNHVTAERFTGS 197

The hypothetical product of *yvf* gene shows similarity to EXOY of *R.meliloti*, an exopolysaccharide production protein. Based on this and on the two predicted transmembrane regions in the homologous *N.gonorrhoeae* sequence, it is predicted that these proteins, or their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 4

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 19>:

```

1  ..AACCATATGG CGATTGTCAT CGACGAATAC GCGCGCACAT CCGGCTTGGT
51  CACCTTTGAA GACATCATCG AGCAAATCGT CCGCGAAATC GAAGACGAGT
101 TTGACGAAGA CGATAGCGCC GACAATATCC ATGCCGTTTC TTCAGACACG
151 TGGCGCATCC ATGCAGCTAC CGAAATCGAA GACATCAACA CCTTCTTCGG
201 CACGGAATAC AGCATCGAAG AAGCCGACAC CATT.GGCGG CCTGGTCATT
251 CAAGAGTTGG GACATCTGCC CGTGGCGCGG GAAAAAGTCC TTATCGGCGG
301 TTTGCAGTTC ACCGTGCGAC GCGCCGACAA CCGCCGCCTG CATACGCTGA
351 TGGCGACCCG CGTGAAGTAA GC..... ACCCG CGTTTCTGCA
15 401 CAGTTTAG

```

This corresponds to amino acid sequence <SEQ ID 20; ORF5>:

```

1  ..NHMAIVIDEY GGTSGLVTFE DIIEQIVGEI EDEFDEDDSA DNIHAVSSDT
51  WRIHAATEIE DINTFFGTEY SIEEADTI XR PGHSRVGTS RARRKSPYRR
101 FAVHRRTRRQ PPPAYADGDP REVS....XR RFCTV*

```

20 Further sequence analysis revealed the complete DNA sequence to be <SEQ ID 21>:

```

1  ATGGACGGCG CACAACCGAA AACGAATTTT TTTGAACGCC TGATTGCCCCG
51  ACTCGCCCCG GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC
101 AGGCGCACGA GCAGGAAGTT TTTGATGCGG ATACGCTTTT AAGATTGGAA
151 AAAGTCTCTG ATTTTCTCGA TTTGGAAGTG CGCGACGCGA TGATTACGGC
25 201 CAGCCGTATG AACGTTTTAA AAGAAAACGA CAGCATCGAG CGCATCACCG
251 CCTACGTTAT CGATACCGCC CATTGCGGCT TCCCCGTCAT CGGCGAAGAC
301 AAAGACGAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTTAACCCC GAGCAGTTCC ACCTCAAATC CATTCTCCGC CCCGCCGTCT
401 TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAGA GTTCCGCGAA
30 451 CAGCGCAACC ATATGGCGAT TGTATCGAC GAATACGGCG GCACATCCGG
501 CTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGC GAAATCGAAG
551 ACGAGTTTGA CGAAGACGAT AGCGCCGACA ATATCCATGC CGTTTCTTCC
601 GAACGCTGGC GCATCCATGC AGCTACCGAA ATCGAAGACA TCAACACCTT
651 CTTCCGCACG GAATACAGCA GCGAAGAAGC CGACACCATT CGGCCTGGTC
35 701 ATTCAAGAGT TGGGACATCT GCCCGTCCGC GCGGAAAAG TCCTTATCGG
751 CGGTTTGCAG TTCACCGTCG CACGCGCCGA CAACCGCCGC CTGCATACCG
801 TGATGGCGAC CCGCGTGAAG TAAGCACCGC CGTTTCTGCA CAGTTTAGGA
851 TGACGGTACG GGCCTTTTCT GTTCAATCC GCCCATCCG CCAAACATAA

```

This corresponds to amino acid sequence <SEQ ID 22; ORF5-1>:

```

40 1  MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLLRLE
51  KVLDFSLEV RDAMITSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMFP EQPHLKSILR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGLVT FEDIEQIVG EIEDEFDEDD SADNIHAVSS
201 ERWRIHAATE IEDINTFFGT EYSSEEADTI RPHGSRVGT S ARARRKSPYR
45 251 RFAVHRRTRR QPPPAYADGD PREVSTAVSA QFRMTVRAFS VSIRPIRQT*

```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 23 >:

```

1  ATGGACGGCG CACAACCGAA AACAAATTTT TTNNAACGCC TGATTGCCCCG
51  ACTCGCCCCG GAACCCGATT CCGCCGAAGA CGTATTGACC CTGTTGCGCC
101 AAGCGCACGA ACAGGAAGTA TTTGATGCGG ATACGCTTTT AAGATTGGAA
50 151 AAAGTCTCTG ATTTTCTGTA TTTGGAAGTG CGCGACGCGA TGATTACGGC
201 CAGCCGTATG AACGTTTTAA AAGAAAACGA CAGCATCGAA CGCATCACCG
251 CCTACGTTAT CGATACCGCC CATTGCGGCT TCCCCGTCAT CGGTGAAGAC
301 AAAGACGAAG TTTTGGGTAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTCAACCCC GAGCAGTTCC ACCTCAAATC GATATTGCGC CCTGCCGTCT

```

This encodes a protein having amino acid sequence <SEQ ID 24; ORF5a>:

```

1 MDGAQPKTNF XXRLIARLAR EPDSAEDVLT LLRQAHEQEV FDADTLRLLE
51 KVLDFSDLEV RDMITRSMN NVLKENDSIE RTIAYVIDTA HSRFPVIGED
101 KQNHGILHA KDLLKYSFM EQFHLKSILR PAVFVPEGKS LTLALLKFRE
151 KDEHMAIVIE EYGGTSGLVT FEDIEIQIVG DIEDFDEDE SADNIHAVSA
201 ERWRIHAATE IEDINAFEGT EYSSEADTI GGXGHSGIGT PARARRKSYX
251 REXAXHXRXR XQPPPAYADG DPREVSSAVS VQFRMTVRAF SVSIRPIRXT
301 *

```

20 The originally-identified partial strain B sequence (ORF5) shows 54.7% identity over a 124aa overlap with ORF5a:

[illegible]

The complete strain B sequence (ORF5-1) and ORF5a show 92.7% identity in 300 aa overlap:

40		10	20	30	40	50	60
	orf5a.pep	MDGAQPKTNFXXRLIARLAREPDSAEDVLTLLRQAHEQEVFDADTLLRLEKVLDFSLEV					
		: : : : : : :					
	orf5-1	MDGAQPKTNFFERLIARLAREPDSAEDVNLNLRQAHEQEVFDADTLLRLEKVLDFSLEV					
		10	20	30	40	50	60
45		70	80	90	100	110	120
	orf5a.pep	RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFP					
		: : : : : : :					
	orf5-1	RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFP					
		70	80	90	100	110	120
50		130	140	150	160	170	180
	orf5a.pep	EQFHLKSILRPVAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIEEQIVG					
		: : : : : : :					
55	orf5-1	EQFHLKSILRPVAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIEEQIVG					
		130	140	150	160	170	180
60		190	200	210	220	230	240
	orf5a.pep	DIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFFGTEYSSEEADTIGGXGHSIGT					
		: : : : : : : :					
	orf5-1	EIEDEFDEDDSDADNIHAVSSERWRIHAATEIEDINTFFGTEYSSEEADTIRP-GHSRVGT					
		190	200	210	220	230	
		250	260	270	280	290	300

orf5a.pep PARARRKSXYRRXAXHXRXRQPPPAYADGDPREVSSAVSQFRMTVRAFSVSIRPIRXT
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
orf5-1 SARARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSTAVSAQFRMTVRAFSVSIRPIRQT
240 250 260 270 280 290

5 Further work identified the a partial DNA sequence in *N.gonorrhoeae* <SEQ ID 25> which encodes a protein having amino acid sequence <SEQ ID 26; ORF5ng>:

	1	MDGAQPKTNF	FERLIARLAR	EPD5AEDVLN	LLRQAHEQEV	FDADTLTRLE
	51	KVLDFAEVL	RDAMITSRM	NVLKENDSIE	RITAYVIDTA	HSRFPVIGED
10	101	KDEVLGTLHA	KDLLKYMENP	EQFHLKSVLR	PAVVFPEGKS	LTLALKEFRE
	151	QRNHMAIVID	EYGGTSGLVT	FEDIIEQIVG	DIEDEFDEDE	SADDIHSVSA
	201	ERWRHHAATE	IEDINAFFGT	EYGSEEDATI	RLRGHSGIGT	PARARRKSPY
	251	RRFAVHRPR	RQPPPAHADG	DPREVSRACP	HRRFCTV*	

Further analysis revealed the complete gonococcal nucleotide sequence <SEQ ID 27> to be:

15	1	ATGGACGGCG	CACAACCGAA	AACAAATTTT	TTTGAACGCC	TGATTGCCCG
	51	ACTCGCCCGC	GAACCCGATT	CCGCCGAAGA	CGTATTAAAC	CTGCTTCGGC
	101	AGGCCGACGA	ACAGGAAGTT	TTTGATGCCG	ACACACTGAC	CCGGCTGGAA
	151	AAAGTATTGG	ACTTTGCCGA	GCTGGAAGTG	CGCGATTCGA	TGATTACCGC
	201	CAGCCGCATG	AACGTATTGA	AAGAAAACGA	CAGCATCGAA	CGCATACCCG
20	251	CCTACGTCAT	CGATACCGCC	CATTGCGGCT	TCCCCGTCAT	CGGCCGAAGAC
	301	AAAGACGAAG	TTTTGGGCAT	TTTGCACGCC	AAGAACCTGC	TCAAATATAT
	351	TTTCAACCCC	GAGCAGTTCC	ACCTGAAATC	CGTCTTGCGC	CTTCCGTTTT
	401	TCGTGCCCGA	AGGCAAATCT	TTGACCGCCC	TTTTAAAAGA	GTTCCGCGAA
	451	CAGCGCAACC	ATATGGCAAT	CGTCATCGAC	GAATACGGCG	GACACGTCGA
25	501	TTTGGTCACC	TTTGAAGACA	TCATCGAGCA	AATCGTCGGT	GGATCTCGAG
	551	ACGAGTTTGA	CGAAGACGAA	AGCGccgacg	acatCCACTC	cgTTTccgCC
	601	GAACGCTGGC	GCATCCacgc	ggctaCCGAA	ATCGAAGaca	TCAACGCCTT
	651	TTTCGGTACG	GAatacggca	gcgaagaagc	cgacaccatc	cggcggctTG
	701	GTCATTcAGG	AATTGGGACA	CCTGCCCGTG	CGCGCGGAAA	AAGTCCTTAt
30	751	cggcgGTTTG	Cagttcacgc	tCGCCCGCGC	CGACAACCGC	CGCCTGCACA
	801	CGCTGATGGC	GACCCGCGTG	AAGTAAGCAG	AGCCTGCCcg	AccgcggttT
	851	CTCGAcAGTT	TAGGatgACG	gtaCGGTCGT	TTTCTGTTTC	AATCCGCCCC
	901	ATCCGCCAAA	CATAA			

This encodes a protein having amino acid sequence <SEQ ID 28; ORF5ng-1>:

35	1	MDGAQPKTNF	FERLIARLAR	EPDSAEDVLN	LLRQAHEQEV	FDADTLTRLE
	51	KVLDFAELEV	RDAMITSRM	NVLKENDSIE	RITAYVIDTA	HSRFPVIGED
	101	KDEVLGILHA	KDLLKYMENF	EQFHLKSVLR	PAVFVPEGKS	LTALLKEFRE
	151	QRNHMAIVID	EYGGTSGLVT	FEDIIEQIVG	DIEDFEDEGS	SADDIHSVSA
	201	ERWRIHAATE	IEDINAFGT	EYGSEADTI	RRLGHSIGIT	PARARRKSPY
40	251	RRFAVHRRPR	RQPPPAHADG	DPREVSRACP	TAVSAQFRMT	VRSFSVSIRP
	301	IROT*				

The originally-identified partial strain B sequence (ORF5) shows 83.1% identity over a 135aa overlap with the partial gonococcal sequence (ORF5ng):

	orf5	NHMAIVIDEYGGTSGLVTFEDIIEQIVGEI	30
45	orf5ng	FHLKSVLRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVGDI	182
	orf5	EDEFDEDDSadNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTIXRPGHSRVGTSA	90
50	orf5ng	EDEFDEDESADDIHVSVAERWRIHAATEIEDINAFFGTEYGSEEADTIRRLGHSGIGTPA	242
	orf5	RARRKSPYRRFAVHRRTRRQPPAYADGDPREVSX---RRFCTV	131
	orf5ng	RARRKSPYRRFAVHRRPRRQPPAHADGDPREVSRACPHRRFCTV	287

The complete strain B and gonococcal sequences (ORF5-1 & ORF5ng-1) show 92.4% identity in 55 304 aa overlap:

orf5ng-1.pep MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAEQEVFDADTLTRLEKVLDFAELEV

15

20

25

30

Computer analysis of these amino acid sequences indicates a putative leader sequence, and identified the following homologies:

Homology with hemolysin homolog TlyC (accession U32716) of *H.influenzae*

ORF5 and TlyC proteins show 58% aa identity in 77 aa overlap (BLASTp).

40

```
ORF5      2      HMAIVIDEYGGTSGLVTFEDIIEQIVGEIEDEFDEDDSDADNIHAVSSDTWRIHAATEIED 61
           HMAIV+DE+G  SGLVT EDI+EQIVG+IEDEFDE++ AD I  +S  T+  + A  T+I+D
TlyC     166     HMAIVVDEFGAVSGLVTTIEDILEQIVGDIEDEFDEEEIAD-IRQLSRHTYAVRALTDIDD 224

ORF5      62      INTFFGTEYSIEEADTI 78
           N  F  T++  EE DTI
TlyC     225     FNAQENTDFDDEEVDTI 241
```

45 ORF5ng-1 also shows significant homology with TlyC:

SCORES Init1: 301 Initn: 419 Opt: 668
Smith-Waterman score: 668; 45.9% identity in 242 aa overlap

```

50      orf5ng-1.pep          MDGAQPKTNFFERLIARLAR-EPDSAEDVLNLLRQAHEQEVEFDADTLTRLEK
                                | ||| :||: : | : |:~::~::~::~::~::~:| :| :|
tlyc_haein MNDEQQNSNQSENTKKPFFQSIFGRFFQGEELKNREELVEVIRDSEQNLDIDONTREMIEG
                   10       20       30       40       50       60

55      orf5ng-1.pep          VLDFAELEVVRDAMITSRMNVLKENDSIERITAYVIDTAHSRFVPIGE--DKDEVILGILH
                               |:::||||| || ||:: ~::::~:: ~::|||~~~~~| ~:::||||
tlyc_haein   VMEIAELRVRDIMIPRSQIIFIEDQQDLNTCLNTIIESAHSRFVIADADDRDNIVGILH
                       70       80       90       100      110      120

60      orf5ng-1.pep          AKDLLKYMF-NPEQFHLKSVLRPVAFVPEGKSLTALLKEFREQRNHMAIVIDEGGTSGL
                               ||||~::~| :||:||||:||||: : ~::|| :| ~::~:||||~::~|||
tlyc_haein   AKDLLKFIREDAEVFDLSLLRPVVIVPESKRVDRLKDFRSERFHMAIVVDFEFGAVSGL
                   110      120      130      140      150      160
```

		130	140	150	160	170	180
5	orf5ng-1.pep	170	180	190	200	210	220
		VT F E D I I E Q I V G D I E D F E D E S A D D I H S V S A E R W R I H A A T E I E D I N A F F G T E Y G S E E A D					
	tlyc_haein	VT I E D I L E Q I V G D I E D F E D E E I A D - I R Q L S R H T Y A V R A L T D I D D F N A Q F N T D F D D E E V D					
		190	200	210	220	230	
10	orf5ng-1.pep	230	240	250	260	270	280
		T I R R L G H S G I G - T P A R A R R K S P Y R R F A V H R R P R R Q P P A H A D G D P R E V S R A C P T A V S A Q F					
	tlyc_haein	T I G L I M Q T F G Y L P K R G E E I I L K N L Q F K V T S A D S R R L I Q L R V T V P D E H L A E M N N V D E K S E					
		240	250	260	270	280	290

15 Homology with a hypothetical secreted protein from *E.coli*:

ORF5a shows homology to a hypothetical secreted protein from *E.coli*:

20 sp|P77392|YBEX_ECOLI HYPOTHETICAL 33.3 KD PROTEIN IN CUTE-ASNB INTERGENIC REGION
>gi|1778577 (U82598) similar to *H. influenzae* [*Escherichia coli*] >gi|1786879
(AE000170) f292; This 292 aa ORF is 23% identical (9 gaps) to 272 residues of an
approx. 440 aa protein YTFL_HAEIN SW: P44717 [*Escherichia coli*] Length = 292

Score = 212 bits (533), Expect = 3e-54

Identities = 112/230 (48%), Positives = 149/230 (64%), Gaps = 3/230 (1%)

25 Query: 2 DGAQPKTNFXRLIARLAR-EPDSAEDVLTLLRQAHEQEVFDADTLLRLEKVLDFSDLEV 60
D K F L+++L EP + +++L L+R + + ++ D DT LE V+D +D V
Sbjct: 10 DTISNKKGFFSLLSQLFHGEPKNRDELLALIRD SGQNDLIEDTRDMLEGVMDIADQRV 69

30 Query: 61 RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYM-FN 119
RD MI RS+M LK N +++ +I++AHSRFPVI EDKD + GIL AKDLL +M +
Sbjct: 70 RDIMIPRSQMITLKRNTLDECLDVIESAHSRFPVISEDKDHIEGILMAKDLLPFMRSD 129

35 Query: 120 PEQFHLKSILRPVAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGGLVT FEDIIEQIV 179
E F + +LR AV VPE K + +LKEFR QR HMAIVIDE+GG SGLVT EDI+E IV
Sbjct: 130 AEAFSMDKVL RQAVVVPESKRVDRLMKEFRSQRVHMAIVIDEFGGVSGGLVTIEDILELIV 189

Query: 180 GDIEDEFDEDESADNIHVA S A E R W R I H A A T E I E D I N A F F G T E Y S S E E A D T 229
G+IEDE+DE++ D +S W + A IED N FGT +S EE DT

Sbjct: 190 GEIEDEYDEEDDID-FRQLSRHTWTVRALASIEDFNEAFGTHFSDEEVD T 238

40 Based on this analysis, including the amino acid homology to the TlyC hemolysin-homologue from
H. influenzae (hemolysins are secreted proteins), it was predicted that the proteins from
N.meningitidis and *N.gonorrhoeae* are secreted and could thus be useful antigens for vaccines or
diagnostics.

ORF5-1 (30.7kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The
45 products of protein expression and purification were analyzed by SDS-PAGE. Figure 2A shows
the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used
to immunise mice, whose sera were used for Western blot analysis (Figure 1B). These experiments
confirm that ORF5-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 5

50 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 29>:

1 ATGCGCGGCG GCAGGCCGGA TTCCGTTACC GTGCAGATTA TCGAAGGTTTC
51 GCGTTTTCG CATATGAGGA AAGTCATCGA CGCAACGCC C GACATCGGAC

5
10
15
20
25
30
35
40
45
50
55

```

101 ACGACACCAA AGGCTGGAGC AATGAAAAAC TGATGGCGGA AGTTGCGCCC
151 GATGCCTTCA GCGGCAATCC TGAAGGGCAG TTTTCCCGG ACAGCTACGA
201 AATCGATGCG GCGGCGAGTG ATTTGCAGAT TTACCAAACC GCCTACAAGG
251 GCGATGCAAC GCCGCTGAA TGAGGGCATG GGAAAGCAGG CAGGACGGGC
301 TGCCTTATAA AAACCTTAT GAAATGCTGA TTATGGCGAr CCTGGTCGAA
351 AAGGAAACAG GGCATGAAGC CGAsCsCGAC CATGTcGCTT CCGTCTTCGT
401 CAACCGCCTG AAAATCGGTA TGCGCCTGCA AACCgAssCG TCCGTGATTT
451 ACGGCATGGG TCGGCATAC AAGGGCAAAA TCCGTAAAGC CGACCTGCGC
501 CGGCACACGC CGTACAACAC CTACACGCGC GGCGGTCTGC CGCCAACCCC
551 GATTGCGCTG CCC..

```

This corresponds to the amino acid sequence <SEQ ID 30; ORF7>:

15
20
25
30
35
40
45
50
55

```

1 MRGGRPDsvT VQIEGSRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP
51 DAFSGNPEGQ FFPDSYEIDA GGSDLQIYQT AYKAMQRRN EAWESRQDGL
101 PYKNPYEMLI MAXLVEKETG HEAXXDHVAS VFNRLKIGM RLQTXSVIY
151 GMGAAYGKI RKADLRRDTP YNTYTRGGLP PTFIALP..

```

Further sequence analysis revealed the complete DNA sequence <SEQ ID 31>:

20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95

```

1 ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTGTGCGG
51 AGCCGTTTTT GCGGCGCTGC TTTTGTTC TAAGGATAAC GGCAGGGCAT
101 ACCGAATCAA AATTGCCAAA AACCAGGTA TTTCGTGCGT CGGCAGGAAA
151 CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGGCGGCGGC
201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGATTGC
251 CTTCCGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCAGG
301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCGCGTT TTTGCGATAT
351 GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGGACACGAC ACCAAAGGCT
401 GGAGCAATGA AAAACTGATG GCGGAAGTTG CGCCGATGC CTTCAGCGGC
451 AATCCTGAAG GGCAGTTTTT CCCCACAGC TACGAAATCG ATGCGGGCGG
501 CAGTGATTG CAGATTACC AAACCGCCTA CAAGGCGATG CAACGCCGCC
551 TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCT
601 TATGAAATGC TGATTATGGC GAGCCTGGTC GAAAAGGAAA CAGGGCATGA
651 AGCCGACCGC GACCATGTGC CTTCCGTCTT CGTCAACCGC CTGAAAATCG
701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGGCGCA
751 TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
801 CACCTACACG CGCGGCGGTC TGCCGCCAAC CCCGATTGCG CTGCCCGGCA
851 AGGCGGCACT CGATGCCGCC GCCATCCGT CCGGCGAAAA ATACCTGTAT
901 TTCGTGTCCA AAATGGACGG CACGGGCTTG AGCCAGTTCA GCCATGATT
951 GACCGAACAC AATGCCGCCG TCCGCAAATA TATTTGAAA AAATAA

```

This corresponds to the amino acid sequence <SEQ ID 32; ORF7-1>:

40
45
50
55
60
65
70
75
80
85
90
95

```

1 MLRKLKWSA VFLTVSAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
51 LAEDRIVFSR HVLTAAYVL GVHNLHTGT YRLPSEVSAW DILQKMRGGR
101 PDSVTVQIEE GSRFSHMRKV IDATPDIGHD TKGWSNEKLM AEVAPDAFSG
151 NPEGQFFPDS YEIDAGGSDL QIYQAYKAM QRRLEAWES RQDGLPYKNP
201 YEMLIMASLV EKETGHEADR DHVASVFNRL KIGMRLQTD PSVIYGMGAA
251 YKGKIRKADL RRDTPYNTYT RGLPPTPIA LPGKAALDAA AHPGSEKYL
301 FVSKMDGTGL SQFSDHLEH NAAVRKYILK K*

```

45 Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein encoded by yceg gene (accession P44270) of *H. influenzae*

ORF7 and yceg proteins show 44% aa identity in 192 aa overlap:

50
55

```

ORF7 1 MRGGRPDsvT VQIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMA-----EVAPDAFSG 55
      + G+ V+ IEG F RK ++ P + K SNE++ A ++ +
yceg 102 LNSGKEVQFNVKWIIEGKTFKDWKRDLENAPHLVQTLKDKSNEEIFALLDLPDIGQNLK 161

ORF7 56 NPEGQFFPDSYEIDAGGSDLQIYQAYKAMQRRLEAWESRQDGLPYKNPYEMLI MAXLV 115
      N EG +PD+Y +DL++ + + + M++ LN+AW R + LP NPYEMLI+A +V
yceg 162 NVEGWLYPDYNTYTPKSTDELLKRSARMKKALNKAWNERDEDLPLANPYEMLILASIV 221

ORF7 116 EKETGHEAXXDHVASVFNRLKIGMRLQTXSVIYGMGAAYKGKIRKADLRRDTPYNTYT 175
      EKETG VASVF+NRLL M+LQT +VIYGMG Y G IRK DL TPYNTY
yceg 222 EKETGIANERAKVASVFINRLKAKMKLQTDPTVIYGMGENYNGNIRKKDLETKTPYNTYV 281

```

-75-

ORF7 176 RGGLPPTPIALP 187
GLPPTPIA+P
yceg 282 IDGLPPTPIAMP 293

The complete length YCEG protein has sequence:

```

5      1 MKKFLIAILL LILILAGVAS FSYYKMTEFV KTPVNVQADE LLTIERGTTS
      51 SKLATLFEQE KLIADGKLLP YLLKLKPELN KIKAGTYSLE NVKTVQDILLD
     101 LLNSGKEVQF NVKWIEGKTF KDWRKDLENA PHLVQTLKDK SNEEIFALLD
     151 LPDIGQNLLEL KNVEGWLYPD TYNYPKSTD LELLKRSAR MKKALNKAWN
     201 ERDEDLPLAN PYEMLILASI VEKETGIANE RAKVASVFIN RLKAKMKLOT
     251 DPTVIYGMGE NYNGNIRKRD LETKTPYNTY VIDGLPPTPI AMPSESSLQA
    301 VANPEKTDYF YFVADGSGGH KPTRNLNEHN KAVQEYLRWY RSQKNAK

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF7 shows 95.2% identity over a 187aa overlap with an ORF (ORF7a) from strain A of *N.*

15 *meningitidis*:

```

                                     10      20      30
      orf7.pep                      MRGGRPDSVTVQIIIEGSRFSHMRKVIDATP
                                     |||
    20      orf7a      AAYVLGVHNRRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIIEGSRFSHMRKVIDATP
                                     70      80      90      100      110      120

                                     40      50      60      70      80      90
      orf7.pep      DIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLN
      |||
    25      orf7a      DIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAMQRRLN
                                     130      140      150      160      170      180

                                     100      110      120      130      140      150
      orf7.pep      EAWESRQDGLPYKNPYEMLIMAXLVEKETGHEAXXDHVASVFVNRLKIGMRLQTXSVIY
      |||
    30      orf7a      EAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFVNRLKIGMRLQTDPSVIY
                                     190      200      210      220      230      240

                                     160      170      180
      orf7.pep      GMGAAYKGKIRKADLRDTPYNTYTRGGLPPTPIALP
      |||
    35      orf7a      GMGAAYKGKIRKADLRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYLYFVSKM
                                     250      260      270      280      290      300

    40      orf7a      DGTGLSQFSHDLTEHNAAVRKYILKKX
                                     310      320      330

```

The complete length ORF7a nucleotide sequence <SEQ ID 33> is:

```

      1 ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTGA CCGTATCGGC
      51 AGCCGTTTTT CCGCGCTGCT TTTTCGTCCC TAAAGACAAC GGCAGGGCAT
    45 101 ACAGGATTAA AATTGCCAAA AACCAGGGTA TTTTCGTCCGT CGGCAGGAAA
     151 CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGGCGGCGGC
     201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGACTGC
     251 CTTCCGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCAGG
     301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCCGCTT TTTTCGATAT
    50 351 GAGGAAAGTC ATCGACGCAA CGCCGACAT CGAACACGAC ACCAAAGGCT
     401 GGAGCAATGA AAAACTGATG GCGGAAGTTG CCCCTGATGC CTTCAGCGGC
     451 AATCCTGAAG GGCAGTTTTT CCGGACAGC TACGAAATCG ATGCGGGCGG
     501 CAGCGATTTA CGGATTTACC AAATCGCCTA CAAGGCGATG CAACGCCGAC
     551 TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCTT
    55 601 TATGAAATGC TGATTATGGC GAGCCTGATC GAAAAGGAAA CAGGGCATGA
     651 AGCCGACCGC GACCATGTCT CTTCCGTCTT CGTCAACCGC CTGAAAATCG
     701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGGGGCA
     751 TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
     801 CACCTACACG CGCGGCGGTC TGCCGCCAAC CCCGATCGCG CTGCGCGGCA
    60 851 AGCGGCACT CGATGCCGCC GCCCATCCGT CCGGTGAAAA ATACCTGTAT
     901 TTCGTGTCCA AAATGGACGG TACGGGCTTG AGCCAGTTCA GCCATGATTT
     951 GACCGAACAC AACCGCGCCG TTCGCAAATA TATTTTGAAA AAATAA

```

This is predicted to encode a protein having amino acid sequence <SEQ ID 34>:

```

      1 MLRKLLKWSA VFLTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
     51 LAEDRIVFSR HVLTAAYVVL GVHNRLHTGT YRLPSEVSAW DILQKMRGGR
    101 PDSVTQIIE GSRFSHMRKV IDATPDIEHD TKGWSNEKLM AEVAPDAFSG
    151 NPEGQFFPDS YEIDAGGSDL RIYQIAYKAM QRRLEAWES RQDGLPYKNP
    201 YEMLIMASLI EKETGHEADR DHVASVFVNR LKIGMRLQTD PSVIYGMGAA
    251 YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGAALDAA AHPSGEKYLY
    301 FVSKMDGTGL SQFSDHDLTEH NAAVRKYILK K*
  
```

A leader peptide is underlined.

10 ORF7a and ORF7-1 show 98.8% identity in 331 aa overlap:

```

                                     10      20      30      40      50      60
    orf7a.pep  MLRKLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAKNQGISSVGRKLAEDRIVFSR
    orf7-1     MLRKLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAKNQGISSVGRKLAEDRIVFSR
    15
                                     10      20      30      40      50      60
    orf7a.pep  HVLTAAYVVLGVHNRLHTGTGYRLPSEVSAWDILQKMRGGRPDSVTQIIEGSRFSHMRKV
    orf7-1     HVLTAAYVVLGVHNRLHTGTGYRLPSEVSAWDILQKMRGGRPDSVTQIIEGSRFSHMRKV
    20
                                     70      80      90     100     110     120
    orf7a.pep  IDATPDIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAM
    orf7-1     IDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAM
    25
                                     70      80      90     100     110     120
    orf7a.pep  IDATPDIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAM
    orf7-1     IDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAM
    30
                                     130     140     150     160     170     180
    orf7a.pep  IDATPDIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAM
    orf7-1     IDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAM
    35
                                     130     140     150     160     170     180
    orf7a.pep  IDATPDIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAM
    orf7-1     IDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAM
    40
                                     190     200     210     220     230     240
    orf7a.pep  QRRLEAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFVNR LKIGMRLQTD
    orf7-1     QRRLEAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFVNR LKIGMRLQTD
    45
                                     190     200     210     220     230     240
    orf7a.pep  QRRLEAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFVNR LKIGMRLQTD
    orf7-1     QRRLEAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFVNR LKIGMRLQTD
    50
                                     250     260     270     280     290     300
    orf7a.pep  PSVIYGMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGAALDAAAHPSGEEKYLY
    orf7-1     PSVIYGMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGAALDAAAHPSGEEKYLY
    55
                                     250     260     270     280     290     300
    orf7a.pep  PSVIYGMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGAALDAAAHPSGEEKYLY
    orf7-1     PSVIYGMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGAALDAAAHPSGEEKYLY
    60
                                     310     320     330
    orf7a.pep  FVSKMDGTGLSQFSDHDLTEHNAAVRKYILKXX
    orf7-1     FVSKMDGTGLSQFSDHDLTEHNAAVRKYILKXX
    65
                                     310     320     330
    orf7a.pep  FVSKMDGTGLSQFSDHDLTEHNAAVRKYILKXX
    orf7-1     FVSKMDGTGLSQFSDHDLTEHNAAVRKYILKXX
  
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF7 shows 94.7% identity over a 187aa overlap with a predicted ORF (ORF7.ng) from *N.*

gonorrhoeae:

```

    50   orf7      MRGGRPDSVTQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQ   60
        orf7ng   MRGGRPDSVTQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQ   60

    55   orf7      FFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWESRQDGLPYKNPYEMLIMASLIEKETG   120
        orf7ng   FFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWAGRQDGLPYKNPYEMLIMASLIEKETG   120

        orf7      HEAXXDHVASVFVNR LKIGMRLQTXSVIYGMGAAYKGKIRKADLRRDTPYNTYTRGGLP   180
        orf7ng   HEADRDHVASVFVNR LKIGMRLQTDPSVIYGMGAAYKGKIRKADLRRDTPYNTYTGGLP   180

        orf7      PTPIALP
  
```

or7ng || ||||
 PTRIALPGKAAMDAAAHPSGEKYLYFVSKMDGTGLSQFSHDLTEHNAAVRKYILKK 236

An ORF7ng nucleotide sequence <SEQ ID 35> is predicted to encode a protein having amino acid sequence <SEQ ID 36>:

5 1 MRGGRPDSVT VQIEGSRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP
 51 DAFSGNPEGQ FFPDSYEIDA GGSDLQIYQT AYKAMQRRLN EAWAGRQDGL
 101 PYKNPYEMLI MASLIEKETG HEADRDHVAS VFNRLKIGM RLQTDPSVIY
 151 GMGAAYKGI RKADLRRDTP YNTYTGGGLP PTRIALPGKA AMDAAAHPSG
 201 EKYLYFVSKM DGTGLSQFSH DLTEHNAAVR KYILKK*

10 Further sequence analysis revealed a partial DNA sequence of ORF7ng <SEQ ID 37>:

 1 ..taccgaatca AGATTGCCAA AAATCAGGGT ATTCGTCGG TCGGCAGGAA
 51 ACTTGCCgaA GACCGCATCG TGTTACAGCAG GCATGTTTTG ACAGCGGCGG
 101 CCTACGTTTT GGGTGTGCAC AACAGGCTGC ATACGGGGAC gTACAGATTG
 151 CCTTCGGAAG TGTCTGCTTG GGATATCTTG CAGAAAATGC GCGGCGGCAG
 15 201 GCCGGATTCC GTTACCGTGC AGATTATCGA AGGTTCCGCT TTTTCGCATA
 251 TGAGGAAAGT CATCGACGCA ACGCCCGACA TCGGACACGA CACCAAAGGC
 301 TGGAGCAATG AAAAATGAT GCGGGAAGTT GCGCCCGATG CCTTCAGCGG
 351 CAATCCTGAA GGGCAGTTTT TTCCCGACAG CTACGAAATC GATGCGGGCG
 401 GCAGCGATTT GCAGATTAC CAAACCGCCT ACAAGGCGAT GCAACGCGCG
 20 451 CTGAACGAGG CATGGGCAGG CAGGCAGGAC GGGCTGCCTT ATAAAAACCC
 501 TTATGAAATG CTGATTATGG CGAGCCTGAT CGAAAAGGAA ACGGGGCATG
 551 AGGCCGACCG CGACCATGTC GCTTCCGTCT TCGTCAACCG CCTGAAAATC
 601 GGTATGCGCC TGCAAACCGA CCCGTCCGTG ATTTACGGCA TGGGTGCGGC
 651 ATACAAGGGC AAAATCCGTA AAGCCGACCT GCGCCGCGAC ACGCCGTACA
 25 701 aCaccTAtac gggcgggggc ttgcccgcga cccggattgc gctgcccggc
 751 Aaggcgggcaa tggatgccgc cgcccacccg tccggcggaAa aatacctgTa
 801 tttcgtgtcc AAAATGGACG GCACGGGCTT GAGCCAGTTC AGCCATGATT
 851 TGACCGAACA CAACGCGGcc gTccGCAAT ATATTTTGAA AAAATAA

This corresponds to the amino acid sequence <SEQ ID 38; ORF7ng-1>:

30 1 ..YRIKIAKNQG ISSVGRKLAE DRIVFSRHVL TAAAYVLGVH NRLHTGTYRL
 51 PSEVSAWDIL QMRGGRPDS VTVQIEGSR FSHMRKVIDA TPDIGHDTKG
 101 WSNEKLMAEV APDAFSGNPE GQFFPDSYEI DAGGSDLIY QTAYKAMQRR
 151 LNEAWAGRQD GLPYKNPYEM LIMASLIEKE TGHEADRDHV ASVFNRLKI
 35 201 GMRLQTDPSV IYGMGAAYKG KIRKADLRD TPYNTYTGGG LPPTRIALPG
 251 KAAMDAAHP SGEKYLYFVS KMDGTGLSQF SHDLTEHNAA VRKYILKK*

ORF7ng-1 and ORF7-1 show 98.0% identity in 298 aa overlap:

	10	20	30	40	50	60
orf7-1.pep	KLLKWSAVFLTVSAAVFAALLFVPKDN	GRAYRIKIAKNQGISSVGRKLAEDRIVFSRHVL				
40 orf7ng-1				YRIKIAKNQGISSVGRKLAEDRIVFSRHVL		
		10	20	30		
	70	80	90	100	110	120
orf7-1.pep	TAAAYVLGVHNRLHTGTYRLPSEVSAWDILOKMRGGRPDSVTVQIEGSRFSHMRKVIDA					
45 orf7ng-1	TAAAYVLGVHNRLHTGTYRLPSEVSAWDILOKMRGGRPDSVTVQIEGSRFSHMRKVIDA					
	40	50	60	70	80	90
	130	140	150	160	170	180
orf7-1.pep	TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLIYQTAYKAMQRR					
50 orf7ng-1	TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLIYQTAYKAMQRR					
	100	110	120	130	140	150
	190	200	210	220	230	240
orf7-1.pep	LNEAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFNRLKIGMRLQTDPSV					
55 orf7ng-1	LNEAWAGRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFNRLKIGMRLQTDPSV					
	160	170	180	190	200	210
	250	260	270	280	290	300
orf7-1.pep	IYGMGAAYKGIKIRKADLRRDTPYNTYTGGGLPPTPIALPGKAALDAAAHPSGEKYLYFVS					

```

      |||
orf7ng-1  IYGMGAAYKKGKIRKADLRDTPYNTYTGGGLPPTRIALPGKAAMDAAAHPSGEKYLYFVS
           220      230      240      250      260      270
5
      310      320      330
orf7-1.pep  KMDGTGLSQFSHDLTEHNAAVRKYILKKX
           |||
orf7ng-1  KMDGTGLSQFSHDLTEHNAAVRKYILKKX
           280      290

```

In addition, ORF7ng-1 shows significant homology with a hypothetical *E.coli* protein:

sp|P28306|YCEG_ECOLI_HYPOTHETICAL_38.2_KD_PROTEIN_IN_PABC-HOLB_INTERGENIC_REGION
gi|1787339 (AE000210) o340; 100% identical to fragment YCEG_ECOLI SW: P28306 but
has 97 additional C-terminal residues [Escherichia coli] Length = 340
Score = 79 (36.2 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57
Identities = 20/87 (22%), Positives = 40/87 (45%)

Query: 10 GISSVGRKLAEDRIVFSRHVLTAAYVLGVHNLHTGTYRLPSEVSAWDILOKMRGGRPD 69
G ++G +L D+I+ V + + GTYR +++ ++L+ + G+
Sbjct: 49 GRLALGEQLYADKIINRPRVFQWLLRIEPLDLSHFAGTYRFTPMQTVREMLKLLSESGKEA 108

Query: 70 SVTVQIIEGSRFSHMRKVIDATPDIGH 96
+++EG R S K + P I H
Sbjct: 109 QFPLRLVEGMRLSDYLLKQLREAPYIKH 135

Score = 438 (200.7 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57
Identities = 84/155 (54%), Positives = 111/155 (71%)

Query: 120 EGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLEAWAGRQDGLPYKNPYEMLIMASLIEK 179
EG F+PD++ A +D+ + + A+K M + ++ AW GR DGLPYK+ +++ MAS+IEK
Sbjct: 158 EGWFWPDTWMTANTTDVALLKRAHKMKVKAVDSAWEGRADGLPYKDKNQLVTMASIEK 217

Query: 180 ETGHEADRDHVASVFNRLKIGMRLQTDPSVIYMGMAAYKKGKIRKADLRDTPYNTYTGG 239
ET ++RD VASVF+NRL+IGMRLQTD+VIYGMG Y GK+ +ADL T YNTYT
Sbjct: 218 ETAVASERDKVASVFINRLRIGMRLQTDPTVIYMGGERYNGKLSRADLETPTAYNTYTIT 277

Query: 240 GLPPTRIALPGKAAMDAAAHPSGEKYLYFVSKMDG 274
GLPP IA PG ++ AAAHP+ YLYFV+ G
Sbjct: 278 GLPPGAIATPGADSLKAAAHPAKTPYLYFVADGKG 312

Based on this analysis, including the fact that the *H.influenzae* YCEG protein possesses a possible leader sequence, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 6

45 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 39>:

```

      1  CGTTTCAAAA TGTTAACTGT GTTGACGGCA ACCTTGATTG CCGGACAGGT
51  ATCTGCCGCC GGAGGCGGTG CGGGGGATAT GAAACAGCCG AAGGAAGTCG
101 GAAAGGTTTT CAGAAAGCAG CAGCGTTACA GCGAGGAAGA AATCAAAAAC
151 GAACGCGCAC GGCTGCGGC AGTGGGCGAG CGGGTTAATC AGATATTTAC
50 201 GTTGCTGGGA GGGGAAACCG CCTTGCAAAA GGGGACGGCG GGAACGGCTC
251 TGGCAACCTA TATGCTGATG TTGGAACGCA CAAATCCCC CGAAGTCGCC
301 GAACGCGCCT TGGAAATGGC CGTGTGCTG AACCGGTTG AACAGGCGGA
351 AATGATTAT CAGAAATGGC GGCAGATTGA GCCTATACCG GGTAAGGCGC
401 AAAACGGGC GGGGTGGCTG CGGAACGTGC TGAGGGAAAG AGGAAATCAG
55 451 CATCTGACG GACGGGAAGA AGTGTGCTG CAGGCGGACG AAGGACAG

```

This corresponds to the amino acid sequence <SEQ ID 40; ORF9>:

```

      1  ..RFKMLTVLTA TLIAGQVSAA GGGAGDMKP KEVGKVRKQ QRYSEEEIKN
51  ERARLAAVGE RVNQIFTLG GETALQKGQA GTALATYMLM LERTKSPEVA
101  ERALEMAVSL NAFEQAEMIV QKWRQIEPIP GKAQKRAGWL RNVLRERGNQ

```

151 HLDGREEVLA QADEGQ

Further sequence analysis revealed the complete DNA sequence <SEQ ID 41>:

	1	ATGTTACCTA	ACCGTTTCAA	AATGTTAACT	GTGTTGACGG	CAACCTTGAT
5	51	TGCCGGACAG	GTATCTGCCG	CCGGAGGCGG	TGCCGGGGAT	ATGAAACAGC
	101	CGAAGGAAGT	CGGAAAGGTT	TTCAGAAAGC	AGCAGCGTTA	CAGCGAGGAA
	151	GAAATCAAAA	ACGAACGCGC	ACGGCTTGCG	CGAGTGGCGC	AGCGGGTTAA
	201	TCAGATATTT	ACGTTGCTGG	GAGGGGAAAC	CGCCTTGCAA	AAGGGGCGAG
10	251	CGGGAACGGC	TCTGGCAACC	TATATGCTGA	TGTTGGAACG	CACAAAAATC
	301	CCCGAAGTCG	CCGAACGCGC	CTTGGAAATG	CGCGTGTGCG	TGAACGCGTT
	351	TGAACAGGCG	GAAATGATTT	ATCAGAAATG	GCGGCAGATT	GAGCCTATAC
	401	CGGGTAAGGC	GCAAAAAACG	GCGGGGTGGC	TGCGGAACGT	GCTGAGGGAA
15	451	AGAGGAAATC	AGCATCTGGA	CGGACTGGAA	GAACTGCTGG	CTCAGGCGGA
	501	CGAAGGACAG	AACCGCAGGG	TGTTTTTATT	TGTGGCACA	CGCCCGGTGC
	551	AACAGGACGG	GTTGGCGCAA	AAAGCATCGA	AAGCGGTTCC	CCGCGCGGCG
	601	TTGAAATATG	AACATCTGCC	CGAAGCGGCG	GTTGCCGATG	TGGTGTTCAG
20	651	CGTACAGGGA	CGCGAAAAGG	AAAAGGCAAT	CGGAGCTTTG	CAGCGTTTGG
	701	CGAAGCTCGA	TACGGAAATA	TGCCCCCA	CTTTAATGAC	GTTGCGTCTG
	751	ACTGCACAGA	AATATCCCGA	AATACTCGAC	GGCTTTTTCC	AGAGACAGAA
	801	CACCCAAAAA	CTTTCGGCCG	TCTGGCAGGA	CAATGGAAAT	ATGAATCTGG
25	851	TTTCCCTGCA	CAGGCTGGAT	GATGCCTATG	CGCGTTTGAA	CGTGTCTGTT
	901	GAACGCAATC	CGAATGCAGA	CCTGTATATT	CAGGCAGCGA	TATTGGCGGC
	951	AAACCGAAAA	GAAGGTGCTT	CCGTATTCGA	CGGCTACGCC	GAAAAGGCAT
	1001	ACGGCAGGGG	CAGCGAGGAA	CAGCGGAGCA	GGGCGGCCTC	AACGGCGGCG
30	1051	ATGATGTATG	CCGACGCGAG	GGATTACGCC	AAAGTCAGGC	AGTGGCTGAA
	1101	AAAAGTATCC	GCGCCGGAAT	ACCTGTTCTG	CAAAGGTGTG	CTGGCGGCTG
	1151	CGGCGGCTGT	CGAGTTGGAC	GGCGGCAGGG	CGGCTTTGCG	CGCATCTGGC
	1201	AGGGTGCGGA	AAC TTCGCA	ACAGCAGGGG	CGGTATTTTA	CGCGAGACAA
35	1251	TTGTGCCAAA	ATACAGATGC	TGCCCCGTG	GAAGCTGCCC	GATAAACGGG
	1301	AGGCTTTGAG	GGGGTTGGAC	AAGATTATCG	AAAAACCGCC	TGCCCGCAGT
	1351	AATACAGAGT	TACAGGCAGA	GGCATTGGTA	CAGCGGTGAG	TGTGTTACGA
	1401	TCGGCTTGGC	AAGCGGAAAA	AAATGATTTT	AGATCTTGAA	AGGGCGGTTT
40	1451	GGCTTGCACC	CGATAACGCT	CAGATTATGA	ATAATCTGGG	CTACAGCCTG
	1501	CTGACCGATT	CCAAACGTTT	GGACGAAGGT	TTGCGCCTGC	TTGACAGCGC
	1551	ATACCAAATC	AACCCGGGAC	ATACCGCTGT	CACACGACG	ATAGGCTGGG
	1601	CGTATTACCT	GAGGAGCGAC	GCGGAAAGCG	CGCTGCGGTA	TCTGCGGTAT
45	1651	TCGTTTGAAG	ACGACCCCGA	GCCCGAAGTT	GCCGCCCAT	TGGGCGAAGT
	1701	GTTGTGGGCA	TGCGGCGAAC	GCGATCAGCG	GGTGTAGCTA	TGGACCGCAG
	1751	CGGCACACCT	TTGGGGAGAC	AAGAAAATAT	GCGCGGAAAC	GCTCAACGTT
	1801	CACGGCATCG	CATTGCCCCA	ACCTTCCCGA	AAACCTCGGA	AATAA

40 This corresponds to the amino acid sequence <SEQ ID 42; ORF9-1>:

	1	MLPNRFKMLT	VLTATLIAGO	VSAAGGGAGD	MKQPKEVGKV	FRKQORYSEE
	51	EIKNERARLA	AVGERVUNQIF	TLLGGETALQ	KGOAGTALAT	YMLMLERTKS
	101	PEVAERALEM	AVSLNAFEQA	EMIYQKWQRI	EIPPGKAQKR	AGWLRNVLR
45	151	RGNQHLDGLE	EVLAQADEGQ	NRRVFLLLAQ	AAVQQDGLAQ	KASKAVRRAA
	201	LKYEHLPEAA	VADVVFVSQG	REKEKAIGAL	ORLAKLDTEI	LPPTLMTLRL
	251	TARKYPEILD	GFFEQTDTQN	LSAVWQEMEI	MNLVLSLHRLD	DAYARLNVLL
	301	ERNPNADLYI	QAAILAANRK	EGASVIDGYA	EKAYGRGTEE	QRSRAALTA
	351	MMYADRRDYA	KVRQWLKKVS	APEYLFDKGV	LAAAAAVELD	GGRAALRGIA
	401	RVRKLPEQQG	RYFTADNLSK	IQMLALSKLP	DKREALRGLD	KIIEKPPAGS
50	451	NTELQAEALV	QRSVVYDRLG	KRKKMISDLE	RAFRLAPDNA	QIMNNGYSL
	501	LTDSKRLEDE	FALLQATAYQI	NPDDTAVNDS	IGWAYYLKGD	AESALPYLRY
	551	SFENDPEPEV	AHLGLVEWLVA	LGERDQAVDV	WTQAAHLTGD	KKIWRETLKR
	601	HGIALPQPSR	KPRK*			

Computer analysis of this amino acid sequence gave the following results:

55 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF9 shows 89.8% identity over a 166aa overlap with an ORF (ORF9a) from strain A of *N.*

meningitidis:

```

                                10      20      30      40      50
60 orf9.pep          RFKMLTVLTATLIAGQVSAAGGGAGDMKPKEVGKVFRKQORYSEEEIKNERARLA
    ||::|::|::|::|::|| |::| | |||||||||||||||||||||
orf9a               MLPARFTILSVLAAALLAGQAYAA--GAADAKPPKEVGKVFRKQORYSEEEIKNERARLA

```

-80-

		10	20	30	40	50
		60	70	80	90	100
5	orf9.pep	AVGERVNIQIF	TLGGETALQK	QAGTALATY	MLMLERTKS	PEVAERALEMAVSLNAFEQA
	orf9a	AVGERVNIQIF	TLGGETALQK	QAGTALATY	MLMLERTKS	PEVAERALEMAVSLNAFEQA
		60	70	80	90	100
		120	130	140	150	160
10	orf9.pep	EMIQKWROIE	PIPGKAQK	RAGWLRNV	LRERGNQHL	DGREEVLAQADEGQ
	orf9a	EMIQKWROIE	PIPGKAQK	RAGWLRNV	LRERGNQHL	DGLEEXLAQADEXQNRVFLLLAQ
		120	130	140	150	160
15	orf9a	AAVQQDGLAQ	KASKAVRRA	ALRYEHLPE	AAVADVFSV	QXREKEKAIGALQRLAKLDEI
		180	190	200	210	220

The complete length ORF9a nucleotide sequence <SEQ ID 43> is:

	1	ATGTTACCCG	CCCGTTTCAC	CATTTTATCT	GTGCTCGCGG	CAGCCCTGCT
	51	TGCCGGGCGAG	GCGTATGCCG	CCGGCGCGGC	GGATGCGAAG	CCGCCGAAGG
20	101	AAGTCGGAAA	GGTTTTCAGA	AAGCAGCAGC	GTTACAGCGA	GGAAGAATC
	151	AAAAACGAAC	GCGCACGGCT	TGCGGCAGTG	GGCGAGCGGG	TTAATCAGAT
	201	ATTTACGTTG	CTGGGANGGG	AAACCGCCTT	GCAAAAGGGG	CAGGCGGGAA
	251	CGGCTCTGGC	AACCTATATG	CTGATGTTGG	AACGCACAAA	ATCCCCCGAA
	301	GTCGCCGAAC	GCGCCTTGGA	AATGGCCGTG	TCNCTGAACG	CGTTTGAACA
25	351	GGCGGAAATG	ATTTATCAGA	AATGGCGGCA	GATTGAGCCT	ATACCGGGTA
	401	AGGCGCAAAA	ACGGGCGGGG	TGGCTGCGGA	ACGTGCTGAG	GGAAAGAGGA
	451	AATCAGCATC	TAGACGGAAT	GGAAGAANTG	CTGGCTCAGG	CGGACGAANG
	501	ACAGAACCGC	AGGGTGTGTT	TATTGTTGGC	ACAAGCCGCC	GTGCAACAGG
	551	ACGGGTGCGC	GCAAAAAGCA	TCGAAAGCGG	TTGCGCCGCG	GGCGTTGAGA
30	601	TATGAACATC	TGCCCCGAAG	GGCGGTTGCC	GATGTGGTGT	TCAGCGTACA
	651	GGNACGCGAA	AAGGAAAAGG	CAATCGGAGC	TTTGCAGCGT	TTGGCGAAGC
	701	TCGATACGGA	AATATTGCCC	CCCACTTTAA	TGACGTTGCG	TCTGACTGCA
	751	CGCAAAATATC	CCGAAATACT	CGACGGCTTT	TTGAGCAGCA	CAGACACCCA
	801	AAACCTTTCG	GCGCTCTGGC	AGGAAATGGA	AATATGAAT	CTGGTTTCCC
35	851	TGCACAGGCT	GGATGATGCC	TATGCGCGTT	TGAACGTGCT	GTTGGAACGC
	901	AATCCGAATG	CAGACCTGTA	TATTCAGGCA	GCGATATTGG	CGGCAAAACG
	951	AAAAGAANGT	GCTTCCGTTA	TCGACGGCTA	CGCCGAAAAG	GCATACGGCA
	1001	GGGGGACGGG	GGAACAGCGG	GGCAGGGCGG	CAATGACGGC	GGCGATGATA
	1051	TATGCCGACC	GAAGGGATTA	CACCAAAGTC	AGGCAGTGGT	TGAAAAAAGT
40	1101	GTCCGCGCCG	GAATACCTGT	TCGACAAAGG	TGTGCTGGCG	GCTGCGGCGG
	1151	CTGTCGAGTT	GGACNGCGGC	AGGGCGGCTT	TGCGGCAGAT	CGGCAGGGTG
	1201	CGGAAACTTC	CCGAACAGCA	GGGGCGGTAT	TTTACGGCAG	ACAATTTGTC
	1251	CAAAATACAG	ATGTTCCGCC	TGTCGAAGCT	GCCCGACAAA	CGGGAGGCTT
	1301	TGAGGGGGTT	GGACAAGATT	ATCGAAAAAC	CGCCTGCCGG	CAGTAATACA
45	1351	GAGTTACAGG	CAGAGGCATT	GGTACAGCGG	TCAGTTGTTT	ACGATCGGCT
	1401	TGGCAAGCGG	AAAAAATGA	TTTCAGATCT	TGAAAGGGCG	TTCAGGCTTG
	1451	CACCCGATAA	CGCTCAGATT	ATGAATAATC	TGGGCTACAG	CCTGCTTCC
	1501	GATTCCAAAC	GTTTGGACGA	AGGCTTCGCC	CTGCTCAGA	CGGCATACCA
	1551	AATCAACCCG	GACGATACCG	CTGTCAACGA	CAGCATAGGC	TGGGCGTATT
50	1601	ACCTGAAANG	CGACGCGGAA	AGCGCGCTGC	CGTATCTGCG	GTATTCTGTT
	1651	GAAAACGACC	CCGAGCCCGA	AGTTGCCGCC	CATTGCGGCG	AAGTGTGTGTG
	1701	GGCATTGGGC	GAACGCGATC	AGGCGGTTGA	CGTATGGACG	CAGGCGGCAC
	1751	ACCTTACGGG	AGACAAGAAA	ATATGGCGGG	AAACGCTCAA	ACGTCACGGC
	1801	ATCGCATTGC	CCCAACCTTC	CCGAAAACCT	CGGAAATAA	

55 This encodes a protein having amino acid sequence <SEQ ID 44>:

	1	MLPARFTILS	VLAAALLAGO	AYAAGAADAK	PPKEVGKVER	KQORYSEEEI
	51	KNERARLAAY	GERVNIQIFTL	LGXETALQKG	QAGTALATYM	LMMLERTKSPE
	101	VAERALEMAV	SLNAFEQAEM	IYQKWROIEP	IPGKAQKQKAG	WLRNVLRERG
	151	NQHLDGLEEX	LAQADEXQNR	RVFLLLAQAA	VQQDGLAQKA	SKAVRRAALR
60	201	YEHLPEAAVA	DVVFVQXRE	KEKAIGALQR	LAKLDEILP	PTLMTLRLTA
	251	RKYPEILDGF	FEQTDTONLS	AVWQEMEIMN	LVSRLRLDDA	YARLNVLLER
	301	NPNADLYIQA	AILAANRKEK	ASVIDGYAEK	AYGRGTGEOR	GRAAMTAAMI
	351	YADRRDYTKV	RQWLKKVSAP	EYLFDKGVLA	AAAAVELDXG	RAALRQIGRV
	401	RKLPEQQGRY	FTADNLSKIQ	MFALSKLPDK	REALRGLDKI	IEKPPAGSNT
65	451	ELQAEALVQR	SVVYDRLGKR	KKMISDLERA	FRLAPDNAQI	MNNLGYSLLS
	501	DSKRLDEGFA	LLQYAYQINP	DDTAVNDSIG	WAYYLKXDAE	SALPYLRYSF
	551	ENDPEPEVAA	HLGEVLWALG	ERDQAVDVWT	QAAHLTGDKK	IWRETLKRHG

601 IALPQPSRKPK RK*

ORF9a and ORF9-1 show 95.3% identity in 614 aa overlap:

5	orf9a.pep	10 20 30 40 50	MLPARFTILSVLAAALLAGQAYAG--AADAKPPKEVGKVKFRKQORYSEEEIKNERARLA
	orf9-1	10 20 30 40 50 60	MLPNRFKMLTVLTATLIAGQVSAAGGGAGDMKQPKVEVGKVKFRKQORYSEEEIKNERARLA
10	orf9a.pep	60 70 80 90 100 110	AVGERVNOIFTLLGXETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
	orf9-1	70 80 90 100 110 120	AVGERVNOIFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
15	orf9a.pep	120 130 140 150 160 170	EMIQKWRQIEPIPGKAQKRAGWLRNVLRRGNQHLDGLEEXLAQADEXQNRVFLLLAQ
	orf9-1	130 140 150 160 170 180	EMIQKWRQIEPIPGKAQKRAGWLRNVLRRGNQHLDGLEEVLAQADEGQNRVFLLLAQ
20	orf9a.pep	180 190 200 210 220 230	AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVVSQVQREKEKAIGALQRLAKLDTEI
	orf9-1	190 200 210 220 230 240	AAVQQDGLAQKASKAVRRAALKYEHLPEAAVADVVSQVQREKEKAIGALQRLAKLDTEI
30	orf9a.pep	240 250 260 270 280 290	LPPTLMTLRLTARKYPEILDGFFEQTDTQNL SAVWQEMEIMNLVSLHRLDDAYARLNVLL
	orf9-1	250 260 270 280 290 300	LPPTLMTLRLTARKYPEILDGFFEQTDTQNL SAVWQEMEIMNLVSLHRLDDAYARLNVLL
35	orf9a.pep	300 310 320 330 340 350	ERNPNADLYIQAAAILAANKEXASVIDGYAEKAYGRGTGEQGRAAMTAAMIYADRRDYT
	orf9-1	310 320 330 340 350 360	ERNPNADLYIQAAAILAANKEGASVIDGYAEKAYGRGTGEQSRRAALTAAMMYADRRDYA
40	orf9a.pep	360 370 380 390 400 410	KVRQWLKKVSAPEYLFDKGVLA AAAA VELDXGRAALRQIGRVKRLPEQQGRYFTADNL SK
	orf9-1	370 380 390 400 410 420	KVRQWLKKVSAPEYLFDKGVLA AAAA VELDGGRAALRQIGRVKRLPEQQGRYFTADNL SK
45	orf9a.pep	420 430 440 450 460 470	IQMFALS KL PDKREALRGLDKIIEKPPAGSNT ELQAEALVQRSVVYDR LGKRKKMISDLE
	orf9-1	430 440 450 460 470 480	IQMLALS KL PDKREALRGLDKIIEKPPAGSNT ELQAEALVQRSVVYDR LGKRKKMISDLE
50	orf9a.pep	480 490 500 510 520 530	RAFRLAPDNAQIMNNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKXD
	orf9-1	490 500 510 520 530 540	RAFRLAPDNAQIMNNLGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD
55	orf9a.pep	540 550 560 570 580 590	AESALPYLRYSFENDPEPEVA AHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWR ETLKR
	orf9-1	550 560 570 580 590 600	AESALPYLRYSFENDPEPEVA AHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWR ETLKR
60	orf9a.pep	600 610	HGIALPQPSRKPRKX
	orf9-1	610	HGIALPQPSRKPRKX

gonorrhoeae:

5	orf9	RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKVEGVKVRKQQRYSSEEIKNERAR	54
	orf9ng	MIMLPARFTILSVLAAALLAGQAYAA--GAADVLPKEVGKVLKRHRRYSSEEIKNERAR	58
10	orf9	LAAVGERVNIQFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	114
	orf9ng	LAAVGERVNRVFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE	118
	orf9	QAEMIIQKWRQIEPIPGKAQKQKAGWLRNVLRERGNQHLDGREEVLAQADEGQ	166
	orf9ng	QAEMIIQKWRQIEPIPGEAQKPAGWLRNVLKEGGNPHLDRLEEVPQSDYVHQPIMIFLL	178

15 The ORF9ng nucleotide sequence <SEQ ID 45> was predicted to encode a protein having including
acid sequence <SEQ ID 46>:

```

1      MIMLPARFTI LSVLAAALLA GQAYAAGAAD VELPKEVGKV LRKHRRYSEE
51     EIKNERARLA AVGERVNRVF TLLGGETALQ KQAGATGALT YMLMLERTKS
101    PEVAERALEM AVSLNAFEQA EMIIYQKWROI EPIPGEAQKP AGWLRNVLKE
20     GGNPHLDRLE EVPAQSDYVH QPMIFLLLVQ AAVOHGGVAQ KPSKVRPAA
201    YNYEVLPETA GADAVFCVQG PQYEKAIQSF PPCGRNPQTE NIAPPFNELF
251    RPTARPISPK LQDRFFRTEP NLAKPFRPPG PEMETYTQTF PRPLTRNNPT

```

Amino acids 1-28 are a putative leader sequence, and 173-189 are predicted to be a transmembrane domain.

25 Further sequence analysis revealed the complete length ORF9ng DNA sequence <SEQ ID 47>>:

	1	ATGTTACCCG	CCCGTTTCAC	TATTTTATCT	GTCCTCGCAG	CAGCCCTGCT
	51	TGCCGGACAG	GCGTATGCTG	CCGGCGCGGC	GGATGTGGAG	CTGCCGAAGG
	101	AAGTCGGAAG	GGTTTAAAG	AAACATCGGC	GTTACAGCGA	GGAGAAATCT
30	151	AAAAACGAAC	CGCGACGGCT	TGCGGCAGTG	GTCACACCGG	TCAACAGGGT
	201	GTTTACGCTG	TTGGGCGGTG	AAACGGCTTT	GCAGAAAGGG	CAGGCGGGAA
	251	CGGCTCTGGC	AACCTATATG	CTGATGTTGG	AACGCACAAA	ATCCCCCGAA
	301	GTCGCCGAAC	GCGCCTTGGA	AATGGCCGTG	TCGCTGAACG	CGTTTGAACA
	351	GGCGGAAATT	ATTTATCAGA	AATGgcggca	catcgagcgt	ataCcggytg
35	401	aggcgcataa	accgGcgggG	tggctgcgga	acgtattgaa	ggaagggGGa
	451	aaTCAGCATC	TGGAcggggt	gaaagaggTG	Ctggcgcaat	cggacgatGT
	501	GCAAAAacgc	aggaTATTTT	TGCTGCTGTT	GCAAGCCGCT	GTGcagcagg
	551	gTGGGGTGGC	TCAAAAAGCA	TCGAAAGCGG	TTGCGCggtgc	GGcgttgaaG
	601	TATGAACATC	TGCCcgaagc	ggcggTTGCC	GATGcggTGT	TCGGCGTACA
40	651	GGGACGCGAA	AAGGAAAagg	caaTCGAAGC	TTTGCAGCGT	TTGGCGAAGC
	701	TCGATACGGA	AATATTGCCC	CCCACTTTAA	TGACGTTGCG	TCGTACTGCA
	751	CGCAATATC	CCGAAATACT	CGACGGCTTT	TTCGAGCAGA	CAGACACCCA
	801	AAACCTTTcG	GCCGTCTGGC	AGGAAATGGA	AATTATGAAT	CTGGTTTCCC
	851	TGCGTAAGCC	GGATGATGCG	TATGCGCGTT	TGAACGTGCT	GTTTGAACAC
45	901	AACCCGAATG	CAAACTGTA	TATTCAGCGC	CGCATATTGG	CGCCAAACCG
	951	AAAAGAAGGT	GCGTCCGTTA	TCGACGGCTA	CGCCGAAAAG	GCATACGGCA
	1001	GGGGGACGGG	GGAACAGCGG	GGCagggcgg	cAATTgacgyc	GGCGATGATA
	1051	TATGCCGACC	GCAGGGATTA	CGCCAAAGTC	AGGCAGTGGT	TGAAAAAAGT
	1101	GTCGCGCGCG	GATACCTGT	TCGACA AAGG	CGGCTGCGCG	GCTGCGGCGG
50	1151	CTGCCGAATT	GGACGGAGGC	CGGGCGGCTT	TGCGGCAGAT	CGGCAGGGTG
	1201	CGGAAACTTC	CCGAACACGA	GGGGCGGTAT	TTTACGGCAG	ACAATTGTCT
	1251	CAAAATACAG	ATGCTCGCCC	TGTCGAAGCT	CGCCGACAAA	CGGGAAGCCC
	1301	TGATCGGGCT	GAAACAACATC	ATCGCCAAAC	TTTCGGCGGC	GGGAAGCACG
	1351	GAACCTTTGG	CGGAAGCATT	GGCACAGCGT	TCCATTATTT	ACGaacAGTT
55	1401	cggCAAACCG	GGAAAAATGA	TTGCCGACCT	tgaAAcCgcg	CTCAAACTTA
	1451	CGGCCGATAA	TGCACAAATT	ATGAATAATC	TGGGCTACAG	CCTGCTTTCC
	1501	GATTCCAAC	GTTTGGACGA	GGGTTTCGCC	CTGCTTCAGA	CGGCATACCA
	1551	AATCAACCCG	GACGATACCG	CCGTTAAACGA	CAGCATAGGC	TGGGCGGTATT
	1601	ACCTGAAAGG	CCAGcgcgaA	AGCGCGCTGC	CGTATCTGcg	gtattcggtt
	1651	qAAAAACGACC	CCGAGCCCGA	AGTTGCCGCC	CATTTGGGCG	AAGTGTGTG

1701 GGCATTGGGC GAACGCGATC AGGCGGTTGA CGTATGGACG CAGGCGGCAC
 1751 ACCTTAGGGG AGACAAGAAA ATATGGCGGG AGACGCTCAA ACGCTACGGA
 1801 ATCGCCTTGC CCGAGCCTTC CCGAAAACCC CGGAAATAA

This encodes a protein having amino acid sequence <SEQ ID 48>:

5 1 MLPARFTILS VLAAALLAGQ AYAAGAADVE LPKEVGKVLK KHRRYSEEEI
 51 KNERARLAAY GERVNRVFTL LGGETALQKG QAGTALATYM LMLERTKSPE
 101 VAERALEMAV SLNAFEQAEM IYQKWRQIEP IPGEAQKPAG WLRNVLKEGG
 151 NQHL DGLKEV LAQSDDVQKR RIFLLVQAA VQGGGVAQKA SKAVRRAALK
 201 YEHLPEAAVA DAVFGVQGRE KEKATEALQR LAKLDEILP PTLMTLRLTA
 10 251 RKYPEILDGF FEQTDTONLS AVWOEMEIMN LVSLRKPDDA YARLNVLEH
 301 NPNANLYIQA AILAA NRKEG ASVIDGYAEK AYGRGTGEOR GRAAMTAAMI
 351 YADRRDYAKV RQWLKVSAP EYLFDKGVLA AAAAELDGG RAALRQIGRV
 401 RKLPEQQGRY FTADNLSKIQ MLALS KLDPK REALIGLNNI IAKLSAAGST
 451 EPLAEALAQR SIIYEQFGKR GKMIADLETA LKLTDPNAQI MNNLGYSLLS
 15 501 DSKRLDEGFA LLQTAYQINP DDTAVNDSIG WAYYLGDAE SALPYLRYSF
 551 ENDPEPEVAA HLGEVLWALG ERDQAVDVWT QAAHLRGDKK IWRETLKRYG
 601 IALPEPSRKP RK*

ORF9ng and ORF9-1 show 88.1% identity in 614 aa overlap:

20	orf9-1.pep	10 20 30 40 50 60	MLPNRFKMLTVLTATLIAGVSAAGGGAGDMKQPKVEVGKVFRRKQRYSEEEIKNERARLA : : : : : : : : : :
	orf9ng-1	10 20 30 40 50	MLPARFTILSVLAAALLAGQAYAAG--AADVELPKEVGKVLKHHRRYSEEEIKNERARLA
25	orf9-1.pep	70 80 90 100 110 120	AVGERVNIQIFTLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA :
	orf9ng-1	60 70 80 90 100 110	AVGERVNRVFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
30	orf9-1.pep	130 140 150 160 170 180	EMIIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLGLEEVLAQADEGQNRVFLLLAQ :
	orf9ng-1	120 130 140 150 160 170	EMIIYQKWRQIEPIPGEAQKPAGWLRNVLKEGGNQHLDGLKEVLAQSDDVQKRRIFLLLVQ
35	orf9-1.pep	190 200 210 220 230 240	AAVQQDGLAQKASKAVRRAALKYEHLPAAVADVFSVQGREKEKAIGALQRLAKLDTEI :
	orf9ng-1	180 190 200 210 220 230	AAVQQGGVAQKASKAVRRAALKYEHLPAAVADVFGVQGREKEKAIEALQRLAKLDTEI
40	orf9-1.pep	250 260 270 280 290 300	LPPTLMTLRLTARKYPEILDGFFEQTDTONLSAVWQEMEIMNLVSLHRLDDAYARLNVLL :
	orf9ng-1	240 250 260 270 280 290	LPPTLMTLRLTARKYPEILDGFFEQTDTONLSAVWQEMEIMNLVSLRKPDDAYARLNVLL
45	orf9-1.pep	310 320 330 340 350 360	ERNPNADLYIQAAILAANRKEGASVIDGYAEKAYGRGTGEQSRRAALTAAMMYADRRDYA : :
	orf9ng-1	300 310 320 330 340 350	EHNPNANLYIQAAILAANRKEGASVIDGYAEKAYGRGTGEQGRRAAMTAAMIYADRRDYA
50	orf9-1.pep	370 380 390 400 410 420	KVRQWLKKVSAPEYLFDKGVLA AAAA AVELDGGRAALRQIGRVKRLPEQQGRYFTADNLSK :
	orf9ng-1	360 370 380 390 400 410	KVRQWLKKVSAPEYLFDKGVLA AAAA AVELDGGRAALRQIGRVKRLPEQQGRYFTADNLSK
55	orf9-1.pep	430 440 450 460 470 480	IQMLALS KLDPKREALRGLDKIIEKPPAGSNTELQAEALVQSVVYDRLGKRKKMISDLE :
	orf9ng-1	420 430 440 450 460 470	IQMLALS KLDPKREALIGLNNIIAKLSAAGSTEPLAEALAQRSIIYEQFGKRGKMIADLE
60		490 500 510 520 530 540	

```

5  orf9-1.pep  RAFRLAPDNAQIMNMLGYSLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD
   orf9ng-1    TALKLTPDNAQIMNMLGYSLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD
   480          490          500          510          520          530

10  orf9-1.pep  AESALPYLRYSFENDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR
   orf9ng-1    AESALPYLRYSFENDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLRGDKKIWRETLKR
   540          550          560          570          580          590

15  orf9-1.pep  HGIALPQPSRKPRKX
   orf9ng-1    YGIALPEPSRKPRKX
   600          610

```

In addition, ORF9ng shows significant homology with a hypothetical protein from *P.aeruginosa*:

```

20  sp|P42810|YHE3_PSEAE HYPOTHETICAL 64.8 KD PROTEIN IN HEMM-HEMA INTERGENIC REGION
   (ORF3)
   >gi|1072999|pir||S49376 hypothetical protein 3 - Pseudomonas aeruginosa >gi|557259
   (X82071) orf3 [Pseudomonas aeruginosa] Length = 576
   Score = 128 bits (318), Expect = 1e-28
   Identities = 138/587 (23%), Positives = 228/587 (38%), Gaps = 125/587 (21%)

25  Query: 67  VFTLLGGETALQKGAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQAEMIYQKWR 126
   +++LL E A Q+ + AL+ Y++ ++T+ P V+ERA +A L A ++A W
   Sbjct: 53  LYSLLVAELAGQRNRFDIALSNYVVQAQKTRDPGVSEAFRIAEYLGADQEQALDTSLLWA 112

30  Query: 127  QIEPIPGEAQKPGAG-----WLRNVLKEGGNQHL DGLKEVLAQSDDVQKRRI 172
   + P +AQ+ A ++ VL G+ H D L A++D + +
   Sbjct: 113  RSAPDNLDQRAAAIQLARAGRYEESMVYMEKVLNGQGDFHDFLALSAAETDPDTRAGL 172

35  Query: 173  FXXXXXXXXXXXXXXXXXKASKAVRRAALKYEHLEAAVADAVFGVQGREKEKAIEALQRLA 232
   ++ KY + + A+ Q ++A+ L+ +
   Sbjct: 173  L-----QSFHLLKKYPNNGQLLFGKALLLQDGRPD EALTLLDENS 214

40  Query: 233  KLDTEILPPTLMTLRLTARK-----YPEILDGFFEQTDTQNL SAVWQEMEIMNLVSLRKP 287
   E+ P L + L + K P + G E D + + + + LV +
   Sbjct: 215  ASRHEVAPLLRSRLQSMKRSDEALPLLKAGIKEHPDDKRVRLAYARL----LVEQNRL 270

45  Query: 288  DDAYARLNVLLLEHNP-----ANLYIQAAI----- 312
   DDA A L++ P+ A +Y++ +
   Sbjct: 271  DDAKAEFAGLVQQFPDDDDDLRFSALVCLQAQAWDEARIYEELVERDSHVDAAHFNLG 330

50  Query: 313  -LAANRKEGASVIDGYAEKAYGRGTGEQGRAAMTAAMIYADRRDYAKVRQWLKKVSAPE 371
   LA +K+ A +D YA+ G G + T ++ A R D A R + P+
   Sbjct: 331  RLAEQKDTARALDEYAQ--VGPGNDFLPAQLRQTDVLLKAGRVDEAAQRLDKARSEQPD 388

55  Query: 372  YLFDKXXXXXXXXXXXXXXXXXRXRIGRVRKLPEQQGRYFTADNLSKIQMLALSCLPKDKR 431
   Y A L I+ ALS +
   Sbjct: 389  Y-----AIQLYLIEAEALSNNDOQE 408

60  Query: 432  EALIGLNNIIAKLSAAGSTEPLAEALAQRSIIYEQFGKRGKMIADLETALKLTPDNAQIM 491
   +A + + + E L L RS++ E+ +M DL + PDNA +
   Sbjct: 409  KAWQAIQEGLKQYP-----EDL-NLLYTRSM LAEKRNDLAQMEKD LRFVIAREPDNAMAL 462

65  Query: 492  NNLGYSLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD AESALPYLRYSE 551
   N LGY+L + R E L+ A+++NPDD A+ DS+GW Y +G A YLR + +
   Sbjct: 463  NALGYTLADRTTRYGEARELILKAHLNPDPAILD SMGWINYRQGLADAERYLRQALQ 522

70  Query: 552  NDPEPEVAHLGEVLWALGERDQAVDVWTQAAHLRGDKKIWRETLKR 598
   P+ EVAHLGEVLWA G + A +W + + D + R T+KR
   Sbjct: 523  RYPDHEVAHLGEVLWAQGRQGDARAIWREYLDKQPDSDVLRRTIKR 569

   gi|2983399 (AE000710) hypothetical protein [Aquifex aeolicus] Length = 545
   Score = 81.5 bits (198), Expect = 1e-14
   Identities = 61/198 (30%), Positives = 98/198 (48%), Gaps = 19/198 (9%)

   Query: 408  GRYFTADNL-SKIQMLALSCLPKDKREALIGLNNIIAKLSAAGSTEPLAEALQ----- 459
   G Y A L K ++LA PDK+E L + +K + + L +

```

Sbjct: 335 GNYEDAKRLIEKAKVLA----PDKKEILFLEADYYSKTKQYDKALEILKKLEKDYPNDSR 390
 Query: 460 ----RSIIYEQFGKRGMIADELTALKLTPDNAQIMNNLGYSLLS--DSKRLDEGFALLQ 513
 +I+Y+ G L A++L P+N N LGYSLL +R++E L++
 5 Sbjct: 391 VYFMEAIVYDNLGDIKNAEKALRKAIELDPENPDYNYLGYSLLLWYGKERVEEAEELIK 450
 Query: 514 TAYQINPDDTAVNDSIGWAYYLKGD AESALPYLRYSF-ENDPEPEVA AHLGEVLWALGER 572
 A + +P++ A DS+GW YYLKGD E A+ YL + E +P V H+G+VL +G +
 10 Sbjct: 451 KALEKDPENPAYIDSMGWVYYLKGDYERAMQYLLKALREAYDDPVVNEHVGDVLLKMGYK 510
 Query: 573 DQAVDVWVQAHLRGDKK 590
 ++A + + +A L + K
 Sbjct: 511 EEARNYYERALKLLEEGK 528

- 15 Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 7

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 49>:

20 1 AACCTCTACG CCGGCCCGCA GACCACATCC GTCATCGCAA ACATCGCCGA
 51 CAACCTGCAA CTGGCCAAAG ACTACGGCAA AGTACACTGG TTCGCCTCCC
 101 CGCTCTTCTG GCTCCTGAAC CAACTGCACA ACATCATCGG CAACTGGGGC
 151 TGGGCGATTA TCGTTTTAAC CATCATCGTC AAAGCCGTAC TGTATCCATT
 201 GACCAACGCC TCTTACCGCT CTATGGCGAA AATGCGTGCC GCCGCACCCA
 251 AACTGCAAGC CATCAAAGAG AAATACGGCG ACGACCGTAT GGCGCAACAA
 25 301 CAGGCGATGA TGCAGCTTTA CACAGACGAG AAAATCAACC CG₂CTGGGGC
 351 GCTGCCTGCC TATGCTGTTG CAAATCCCCG TCTTCATCGG ATTGTATTGG
 401 GCATTGTTTC CCTCCGTAGA ATTGCGCCAG GCACCTTGCG TGGGTTGGAT
 451 TACCGACCTC AGCCGCGCCG ACCCCTACTA CATCCTGCCC ATCATTATGG
 501 CGGCAACGAT GTTCGCCCAA ACTTATCTGA ACCCGCCGCC GACCGACCCG
 30 551 ATGcagGCGA AAATGATGAA AATCATGCCG TTGGTTTTCT CsGwCrTgTT
 601 CTTCTTCTTC CCTGCCGGks TGGTATTGTA CTGGGTAGTC AACACCTCC
 651 TGACCATCGC CCAGCAATGG CACATCAACC GCAGCATCGA AAAACAACGC
 701 GCCCAAGGCG AAGTCGTTTC CTAA

This corresponds to the amino acid sequence <SEQ ID 50; ORF11>:

35 1 ..NLYAGPQTS VIANIADNLQ LAKDYGKVHW FASPLFWLLN QLHNIIGNWG
 51 WAIIVLTIIV KAVLYPLTNA SYRSMAMRA AAPKLQAIKE KYGDDRMAQQ
 101 QAMMQLYTDE KINPLGGCLP MLLQIPVFIG LYWALFASVE LRQAPWLGIW
 151 TDLSRADPYY ILPIIMAATM FAQTYLNPPP TDPMQAKMMK IMPLVFSXFX
 201 FFFPAGXVLY WVVNNLLTIA QQWHINRSIE KQRAQGEVVS *

- 40 Further sequence analysis revealed the complete DNA sequence <SEQ ID 51>:

1 ATGGATTTTA AAAGACTCAC GGCCTTTTTT GCCATCGCGC TGGTGATTAT
 51 GATCGGCTGG GAAAAGATGT TCCCCACTCC GAAGCCAGTC CCCGCGCCCC
 101 AACAGGCAGC ACAACAACAG GCCGTAACCG CTTCCGCCGA AGCCGCGCTC
 151 GCGCCCGCAA CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGTCAT
 45 201 TGATGAAAAA AGCGGCGACC TGCGCCGGCT GACCTGCTC AAATACAAAG
 251 CAACCGGCGA CGAAAATAAA CCGTTCATCC TGTTTGGCGA CGGCAAAGAA
 301 TACACCTACG TCGCCCAATC CGAACTTTTG GACGCGCAGG GCAACAACAT
 351 TCTAAAAGGC ATCGGCTTTA GCGCACCGAA AAAACAGTAC AGCTTGGAAG
 401 GCGACAAAGT TGAAGTCCGC CTGAGCGCGC CTGAAACAGC CGGTCTGAAA
 50 451 ATCGACAAAG TTTATACTTT CACCAAAGGC AGCTATCTGG TCAACGTCCG
 501 CTTGACATC GCCAACGGCA GCGGTCAAAC CGCCAACCTG AGCGCGGACT
 551 ACCGCATCGT CCGCGACCAC AGCGAAACCG AGGGTCAAGG TTACTTTACC
 601 CACTCTTACG TCGGCCCTGT TGTTTATACC CCTGAAGGCA ACTTCCAAAA
 651 AGTCAGCTTT TCCGACTTGG ACGACGATGC CAAATCCGGC AAATCCGAGG
 55 701 CCGAATACAT CCGCAAACCC CCGACCGGCT GGCTCGGCAT GATTGAACAC
 751 CACTTCATGT CCACCTGGAT TCTCCAACCT AAAGGCAGAC AAAGCGTTTG
 801 CGCCGCAGGC GAGTGCAACA TCGACATCAA ACGCCGCAAC GACAAGCTGT
 851 ACAGCACCAG GTCAGCGTG CTTTAGCCG CCATCCAAAA CGGCGCGAAA
 901 GCCGAAGCCT CCATCAACCT CTACGCCGGC CCGCAGACCA CATCCGTCAT
 60 951 CGCAAACATC GCGACAACC TGCAACTGGC CAAAGACTAC GGCAAAGTAC

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160

1001 ACTGGTTCGC CTCCCCGCTC TTCTGGCTCC TGAACCAACT GCACAACATC
1051 ATCGGCAACT GGGGCTGGGC GATTATCGTT TTAACCATCA TCGTCAAAGC
1101 CGTACTGTAT CCATTGACCA ACGCCTCTTA CCGCTCTATG GCGAAAATGC
1151 GTGCCGCCGC ACCCAAACCTG CAAGCCATCA AAGAGAAATA CGGCGACGAC
1201 CGTATGGCGC AACAAACAGGC GATGATGCAG CTTTACACAG ACGAGAAAAT
1251 CAACCCGCTG GGGCGCTGCC TGCCTATGCT GTTGCAAATC CCCGTCTTCA
1301 TCGGATTGTA TTGGGCATTG TTCGCCTCCG TAGAATTGCG CCAGGCACCT
1351 TGGCTGGGTG GGATTACCGA CCTCAGCCGC GCCGACCCCT ACTACATCCT
1401 GCCCATCATT ATGGCGGCAA CGATGTTTCG CCAAATTAT CTGAACCCGC
1451 CGCCGACCGA CCCGATGCAG GCGAAAATGA TGAATATCAT GCCGTGGGT
1501 TTCTCCGTCA TGTCTCTCTT CTTCCCTGCC GGTCTGGTAT TGTACTGGGT
1551 AGTCAACAAC CTCCTGACCA TCGCCAGCA ATGGCACATC AACCGCAGCA
1601 TCGAAAAACA ACGCGCCCAA GCGGAAGTCG TTTCTCTAA

This corresponds to the amino acid sequence <SEQ ID 52; ORF11-1>:

15
20
25
30
35
40
45
50

1 MDFKRLTAFF AIALVIMIGW EKMFTPKPV PAPQAAQQQ AVTASAEAL
51 APATPITVTT DTVQAVIDEK SGDRLRLTLL KYKATGDENK PFILFGDGKE
101 YTYVAQSELL DAOGNNILKG IGFSAPKKQY SLEGDKVEVR LSAPETRGLK
151 IDKVTFTKG SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT
201 HSYVGPVVYT PEGNFQKVSF SLDLDDAKSG KSEAERYIKT PTGWLGMIEH
251 HFMSTWILQP KGRQSVCAAG ECNIDIKRRN DKLYSTSVSV PLAAIQNGAK
301 AEASINLYAG PQTTSVIANI ADNLQAKDY GKVHWFASPL FWLLNQLHNI
351 IGNWGWAIIV LTIIVKAVLY PLTNASYRSM AKMRAAAPKL QAIKEYGDD
401 RMAQQQAMMO LYTDEKINPL GGCLPMLLQI PVFIGLYWAL FASVELRQAP
451 WLGWITDLR ADPPYILPII MAATMFAQTY LNPPPTDPMQ AKMMKIMPLV
501 FVMFFFFFA GLVLVWVNN LLTIAQQWHI NRSIEKQRAQ GEVVS*

Computer analysis of this amino acid sequence gave the following results:

Homology with a 60kDa inner-membrane protein (accession P25754) of *Pseudomonas putida*

ORF11 and the 60kDa protein show 58% aa identity in 229 aa overlap (BLASTp).

30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000

ORF11 2 LYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLTIIVK 61
LYAGP+ S + ++ L+L DYG + + A P+FWLL +H+++GNWGW+IIVLT+++K
60K 324 LYAGPKIQSKLKELESPGLELTVDYGFLWFIAPQIFWLLQHIHSLGNWGWSIIVLTMLIK 383
ORF11 62 AVLYPLTNASYRSMAMRAAAPKLQAIKEYGDDRXXXXXXXXXXLYTDEKINPLGGCLPM 121
+ +PL+ ASYRSM+MRA APKL A+KE++GDDR LY EKNPLGGCLP+
60K 384 GLFFPLSAASYRSMARMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPI 443
ORF11 122 LLQIPVFIGLYWALFASVELRQAPWLGWITDLRADPPYILPII MAATMFAQTYLNPPPT 181
L+Q+PVF+ LYW L SVE+RQAPW+ WITDLS DP++ILPIIM ATMF Q LNP P
60K 444 LVQMPVFLALYVWLVLESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIQQRNPTTP 503
ORF11 182 DPMQAKMMKIMPLVXXXXXXXXXVAGXVLYWVNNLLTIAQQWHINRSIE 230
DPMQAK+MK+MP++ PAG VLYWVNN L+I+QQW+I R IE
60K 504 DPMQAKVMKMPPIIFTFFFLWFPAGLVLYWVNNCLSSISQQWYITRRIE 552

45 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF11 shows 97.9% identity over a 240aa overlap with an ORF (ORF11a) from strain A of *N.*

meningitidis:

50
55

orf11.pep NLYAGPQTTSVIANIADNLQAKDYGKVHW
|||||
orf11a IKRRNDKLYSTSVSVPLAAIQNGAKSXASINLYAGPQTTSVIANIADNLQXKDYGKVHW
280 290 300 310 320 330
40 50 60 70 80 90
orf11.pep FASPLFWLLNQLHNIIGNWGWAIIVLTIIVKAVLYPLTNASYRSMAMRAAAPKLQAIKE
|||||
orf11a FASPLFWLLNQLHNIIGNWGWAIIVLTIIVKAVLYPLTNASYRSMAMRAAAPKLQAIKE
340 350 360 370 380 390

-87-

		100	110	120	130	140	150
	orf11.pep	KYGDDRMAQQQAMMQLYTD	DEKINPLGGCLPMLLQIPV	FIGLYWALFASVELRQAPW	LGI		
5	orf11a	KYGDDRMAQQQAMMQLYTD	DEKINPLGGCLPMLLQIPV	FIGLYWALFASVELRQAPW	LGI		
		400	410	420	430	440	450
		160	170	180	190	200	210
	orf11.pep	TDLSRADPYIILPII	MAATMFAQTYLNPPPTD	PMQAKMMKIMPLVFSXX	FFFPAGXVLY		
10	orf11a	TDLSRADPYIILPII	MAATMFAQTYLNPPPTD	PMQAKMMKIMPLVFSXX	FFFPAGXVLY		
		460	470	480	490	500	510
		220	230	240			
15	orf11.pep	WVNNLLTIAQQWHINRS	IEKQRAQGEVVSX				
	orf11a	WVNNLLTIAQQWHINRS	IEKQRAQGEVVSX				
		520	530	540			

The complete length ORF11a nucleotide sequence <SEQ ID 53> is:

	1	ANGGATT	TAAAGACTCAC	NGNGTTT	TTCGCCATCGCAC	TGGTGATTAT
20	51	GATCGGATNG	NAAANGATGT	TCCCCACTCC	GAAGCCCGTC	CCC
	101	AACAGACGGC	ACAACAACAG	GCCGTAANCG	CTTCCGCGCA	AGCCGCGCTC
	151	GCGCCCGNAN	CGCCGATTAC	CGTAACGACC	GACACGGTTC	AAGCCGTCAT
	201	TGATGAAAAA	AGCGGCGACC	TGCGCCGCGT	GACCCCTGCTC	AAATACAAAG
	251	CAACCGGCGA	CNAAAATAAA	CCGTTCATCC	TGTTTGGCGA	CGGCAANAA
25	301	TACACCTACN	TCGCCANTC	CGAACTTTG	GACGCGCAGG	GCAACAACAT
	351	TCTAAAAGGC	ATCGGCTTTA	GCGCACCGAA	AAAACAGTAC	AGCTTGGAAG
	401	GCGACAAAGT	TGAAGTCCGC	CTGAGCGCAC	CTGAAACACG	CGGTCTGAAA
	451	ATCGACAAAG	TTTATACTTT	CACCAAAGGC	AGCTATCTGG	TCAACGTC
	501	CTTCGACATC	GCCAAACGGCA	GCGGTCAAAC	CGCCAACCTG	AGCGCGGACT
30	551	ACCGCATCGT	CCGCGACCAC	AGCGAACC	CGAGGTCAAGG	CTACTTTACC
	601	CACTCTTACG	TCGGCCCTGT	TGTTTATACC	CCTGAAGGCA	ACTTCCAAAA
	651	AGTCAGCTTC	TCCGACTTGG	ACGACGATGC	CAANTCCG	GNAAATCCGAGG
	701	CCGAATACAT	CCGCAAAACC	CNGACCGGCT	GGCTCGGCAT	GATTGAACAC
	751	CACTTCATGT	CCACTGGGAT	CCTCCAACCC	AAAGGCGGAC	AAAGCGTTTG
35	801	CGCCGCTGGC	GACTGCNGTA	TNGACATCAA	ACGCCGCAAC	GACAAGCTGT
	851	ACAGCACCGA	CGTCAGCGTG	CCTTTAGCCG	CTATCCAAAA	CGGTGCGAAA
	901	TCCNAAGCCT	CCATCAACCT	CTACGCCGGC	CCACAGACCA	CATCNGTTAT
	951	CGCAAAATC	GCCGACAACC	TGCAACTG	GNCAAAGACTAC	GGCAAAAGTAC
	1001	ACTGGTTCGC	CTCCCCCTC	TTTTGGCTTT	TGAACCAACT	GCACAACATC
40	1051	ATCGGCAACT	GGGGCTGGGC	GATTATCGTT	TTAACCATCA	TCGTCAAAGC
	1101	CGTACTGTAT	CCATTGACCA	ACGCTCTTA	CCGTTCGATG	GCGAAAATGC
	1151	GTGCCGCGC	GCCCAAACTG	CAAGCCATCA	AAGAGAAATA	CGGCGACGAC
	1201	CGTATGGCGC	AGCAACAAGC	CATGATGCAG	CTTTACACAG	ACGAGAAAAAT
	1251	CAACCGCTG	GGCGGCTGCC	TGCCTATGCT	GTTGCAATC	CCCGTCTTCA
45	1301	TCGGATTGTA	TTGGGCATTG	TTCGCCTCCG	TAGAATTGCG	CCAGGCACCT
	1351	TGGCTGGGTT	GGATTACCGA	CCTCAGCCGC	GCCGACCCNT	ACTACATCCT
	1401	GCCCATCATT	ATGGCGGCAA	CGATGTTCCG	CCAAACCTAT	CTGAACCCGC
	1451	CGCCGACCGA	CCCGATGCAG	GCGAAAATGA	TGAAAATCAT	GCCTTTGGTT
	1501	NTNTCNNNNA	NGTTCCTC	CTTCCCTGCC	GGTCTGGTAT	TGTA
50	1551	GATCAACAAC	CTCCTGACCA	TCGCCAGCA	ATGGCACATC	AACCGCAGCA
	1601	TCGAAAAACA	ACGCGCCCAA	GGCGAAGTCG	TTTCTTAA	

This encodes a protein having amino acid sequence <SEQ ID 54>:

	1	XDFKRLTXFF	AIALVIMIGX	XXMFPTPKPV	PAPOQTAQQQ	AVXASAEAL
55	51	APXXPITVTT	DTVQAVIDEK	SGDLRLTL	LKYKATGDXNK	PFILFGDGKX
	101	YTYXAXSELL	DAQGNILKG	IGFSAPKKQY	SLEGDKVEVR	LSAPETRGLK
	151	IDKVYFTFKG	SYLVNVREDI	ANGSGQTANL	SADYRIVRDH	SEPEGQGYFT
	201	HSYVGPVVYT	PEGNFQKVSF	SDLDDAXSG	KSEAEYIRKT	XTGWLGMIEH
	251	HFMSTWILQP	KGGQSVCAAG	DCXXDIKRRN	DKLYSTSVSV	PLAAIQNGAK
	301	SXASINLYAG	PQTTSVIANI	ADNLQLXKDY	GKVHWFASPL	FWLLNLQHLNI
60	351	IGNWGWAIIV	LTIIIVKAVLY	PLTNASYRSM	AKMRAAPKL	QAIKEYGDD
	401	RMAQQQAMMQ	LYTDEKINPL	GGCLPMLLQI	PVFIGLYWAL	FASVELRQAP
	451	WLGWITDLSR	ADPYIILPII	MAATMFAQTY	LNPPPTDPMQ	AKMMKIMPLV
	501	XSXXFFXFFA	GLVLYWVINN	LLTIAQQWHI	NRSIEKQRAQ	GEVVS*

ORF11a and ORF11-1 show 95.2% identity in 544 aa overlap:

65	10	20	30	40	50	60
----	----	----	----	----	----	----

5	orf11a.pep	XDFKRLTXFFAIALVIMIGXXMFEPTPKPVPAPOQTAQQQAVXASAEALAPXXPITVTT
	orf11-1	MDFKRLTAFFAIALVIMIGWEKMEFTPKPVPAPOQAAQQQAVTASAEALAPATPITVTT
10	orf11a.pep	DTVQAVIDEKSGDLRRLTLKLYKATGDXNKPFFILFGDGKXYTYXAXSELLDAQNNILKG
	orf11-1	DTVQAVIDEKSGDLRRLTLKLYKATGDENKPFILFGDGKEYTYVAQSELLDAQNNILKG
15	orf11a.pep	IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYTFTKGSYLVNVRFDIANGSGQTANL
	orf11-1	IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYTFTKGSYLVNVRFDIANGSGQTANL
20	orf11a.pep	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAXSGKSEAEYIRKT
	orf11-1	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAXSGKSEAEYIRKT
25	orf11a.pep	XTGWLGMIEHHFMSTWILQPKGGQSVCAAGDCXXDIKRRNDKLYSTSVSVPLAAIQNGAK
	orf11-1	PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSVPLAAIQNGAK
30	orf11a.pep	SXASINLYAGPQTTSVIANIADNLQKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV
	orf11-1	AEASINLYAGPQTTSVIANIADNLQKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV
35	orf11a.pep	LTIIIVKAVLYPLTNASYRSMKMRAPKLAKEKYGDDRMAQQQAMMQLYTDEKINPL
	orf11-1	LTIIIVKAVLYPLTNASYRSMKMRAPKLAKEKYGDDRMAQQQAMMQLYTDEKINPL
40	orf11a.pep	GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPIIIMATMFAQTY
	orf11-1	GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPIIIMATMFAQTY
45	orf11a.pep	LNPPPTDPMQAKMMKIMPLVXSXXFFXFPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ
	orf11-1	LNPPPTDPMQAKMMKIMPLVFSVMFFFPPAGLVLYWVNNLLTIAQQWHINRSIEKQRAQ
50	orf11a.pep	GEVVSX
	orf11-1	GEVVSX

60 Homology with a predicted ORF from *N.gonorrhoeae*

ORF11 shows 96.3% identity over a 240aa overlap with a predicted ORF (ORF11.ng) from *N. gonorrhoeae*:

65	Orf11	NLYAGPQTTSVIANIADNLQKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT	57
	orf11ng	MAVNLYAGPQTTSVIANIADNLQKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT	60

-89-

	orf11	IIVKAVLYPLTNASYRSMKMRAPKLOAIKEKYGDDRMAMQOLYTDEKINPLGG	117
	orf11ng	IIVKAVLYPLTNASYRSMKMRAPPELQTIKEKYGDDRMAMQOLFEEINPLGG	120
5	orf11	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIIMAAATMFAQTYLN	177
	orf11ng	CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIIMAAATMFAQTYLN	180
10	orf11	PPPTDPMQAKMMKIMPLVFSXXFFFFPAGXVLYVWVNNLLTIAQQWHINRSIEKQRAQGE	237
	orf11ng	PPPTDPMQAKMMKIMPLVFSVMFFFFPAGLVLYVWVNNLLTIAQQWHINRSIEKQRAQGE	240
15	orf11	VVS 240	
	orf11ng	VVS 243	

An ORF11ng nucleotide sequence <SEQ ID 55> was predicted to encode a protein having amino acid sequence <SEQ ID 56>:

	1	MAVNLYAGPQ	TTSVIANIAD	NLQAKDYGK	VHWFASPLFW	LLNQLHNIIG
	51	NWGWAIIVLT	IIVKAVLYPL	TNASYRSMK	MRAAPELQT	IKKEYGDDRM
20	101	AQQQAMQLF	EDEEINPLGG	CLPMLLQIPV	FIGLYWALFA	SVELRQAPWL
	151	GWITDLSRAD	PYYILPIIMA	ATMFAQTYLN	PPPTDPMQAK	MMKIMPLVFS
	201	VMFFFFPAGL	VLYWVNNLL	TIAQQWHINR	SIEKQRAQGE	VVS*

Further sequence analysis revealed the complete gonococcal DNA sequence <SEQ ID 57> to be:

	1	ATGGATTTTA	AAAGACTCAC	GGCGTTTTTC	GCCATCGCGC	TGGTGATTAT
	51	GATCGGCTGG	GAAAAAATGT	TCCCCACCCC	GAAACCCGTC	CCCGCGCCCC
	101	AACAGGCGGC	ACAAAAACAG	GCAGCAACCG	CTTCCGCCGA	AGCCGCGCTC
	151	GCGCCCGCAA	CGCCGATTAC	CGTAACGACC	GACACGGTTC	AAGCCGTTAT
	201	TGATGAAAAA	AGTGGCGACC	TGCGCCGGCT	GACCTGCTC	AAATACAAAG
	251	CAACCGGCGA	CGAAAAACAA	CCGTTCGTCC	TGTTTGCGCA	CGGCAAAAGAA
30	301	TACACCTACG	TCGCCCAATC	CGAACTTTTG	GACGCGCAGG	GCAACAACAT
	351	TCTGAAAGGC	ATCGGCTTTA	GCGCACCAGG	AAAACAGTAC	ACCCTCAACG
	401	GCGACACAGT	CGAAGTCCGC	CTGAGCGCGC	CCGAAACCAA	CGGACTGAAA
	451	ATCGACAAAG	TCTATACCTT	TACCAAAGAC	AGCTATCTGG	TCAACGTCCG
	501	CTTCGACATC	GCCAACGGCA	GCGGTCAAAC	CGCCAACCTG	AGCGCGGACT
35	551	ACCGCATCGT	CCGCGACCCG	AGCGAACCCG	AGGGTCAAGG	CTACTTTACC
	601	CACTCTTACG	TCGGCCCTGT	TGTTTATACC	CCTGAAGGCA	ACTTCCAAAA
	651	AGTCAGCTTC	TCCgacTTgg	acgACGATGC	gaaaTccggc	aaATccgagg
	701	ccgaatacat	CCGCAAAACC	ccgaccgggt	ggctcggcat	gattgaacac
	751	cacttcacgt	ccacctggat	cctccAAcct	aaaggcggcc	aaaacgtttg
40	801	cgcccaggga	gactgcccga	tcgacattaa	aCgcccgaac	gacaagctgt
	851	acagcgcaag	cgtcagcggt	cctttaaccg	ctatcccaac	ccggggggcca
	901	aaaccgaaaa	tgggcgTCAA	CCTGTATGCC	GGTCCGCAAA	CCACATCCGT
	951	TATCGCAAAC	ATCGCcgaca	ACCTGCAACT	GGCAAAGAC	TACGGTAAAG
45	1001	TACACTGGTT	CGCATCGCCG	CTCTTCTGGC	TCCTGAACCA	ACTGCACAAC
	1051	ATTATCGGCA	ACTGGGGCTG	GGCAATCGTC	GTTTTGACCA	TCATCGTCAA
	1101	AGCCGTACTG	TATCCATTGA	CCAACGcctc	ctACCGTTTC	ATGGCGAAAA
	1151	TGCGTGccgc	cgcacCcaaA	CTGCAGACCA	TCAAAGAAAA	ATAcgCGGAC
	1201	GACCGTATGG	CGCAACAGCA	AGCGATGATG	CAGCTTTACA	AAgacgAGAA
	1251	AATCAACCCG	CTGGGCGGCT	GTctgcctat	gctgttgCAA	ATCCCCGTCT
50	1301	TCATCGGCTT	GTACTGGGCA	TTGTTGCGCT	CCGTAGAATT	GCGCCAGGCA
	1351	CCTTGGCTGG	GCTGGATTAC	CGACCTCAGC	CGCGCCGACC	CCTACTACAT
	1401	CCTGCCCATC	ATTATGGCGG	CAACGATGTT	CGCCCAAACC	TATCTGAACC
	1451	CGCCGCGGAC	CGACCCGATG	CAGGCGAAAA	TGATGAAAAT	CATGCCGTTG
	1501	GTTTCTCCG	TCATGTTCTT	CTTCTTCCCT	GCCGGTTTGG	TTCTCTACTG
55	1551	GGTGGTCAAC	AACCTCCTGA	CCATCGCCCA	GCAGTGGCAC	ATCAACCGCA
	1601	GCATCGAAAA	ACAACGCGCC	CAAGGCGAAG	TCGTTTCCTA	A

This encodes a protein having amino acid sequence <SEQ ID 58; ORF11ng-1>:

	1	MDFKRLTAFF	AIALVIMIGW	EKMFPPTPKPV	PAPQAAQKQ	AATASAEAL
	51	APATPITVTT	DTVQAVIDEK	SGDLRLRLTL	KYKATGDENK	PFVLFQDGKE
60	101	YTYVAQSELL	DAQNNILKG	IGFSAPKKQY	TLNGDTVEVR	LSAPETNGLK
	151	IDKVYTFETK	SYLVNVRFDI	ANGSGQTANL	SADYRIVRDH	SEPEGQGYFT
	201	HSYVGPVYVT	PEGNFQKVSF	SDLDDDAKSG	KSEAEYIRKT	PTGWLGMIEH
	251	HFMSTWILQP	KGGQNVCAQG	DCRIDIKRRN	DKLYSASVSV	PLTAIPTRGP
	301	KPKMAVNLYA	GPQTTSVIAN	IADNLQAKD	YGRVHWFASP	LFWLLNQLHN

-90-

351 IIGNWGWAIV VLTIIIVKAVL YPLTNASYRS MAKMRRAAPK LQTIKEYGD
 401 DRMAQQQAMM QLYKDEKINP LGGCLPMLLQ IPVFIGLYWA LEASVELRQA
 451 PWLGWITDLS RADPYIILPI IMAATMFAQT YLNPPPTDPM QAKMMKIMPL
 501 VFSVMFFFEF AGLVLVWVN NLLTIAQQWH INRSIEKQRA QGEVVS*

5 ORF11ng-1 and ORF11-1 shown 95.1% identity in 546 aa overlap:

		10	20	30	40	50	60
	orf11ng-1.pep	MDFKRLTAFFAIALVIMIGWEKMEFTPKPVPAPQQAQQAATASAEALAPATPITVTT					
	orf11-1	MDFKRLTAFFAIALVIMIGWEKMEFTPKPVPAPQQAQQAATASAEALAPATPITVTT					
10		10	20	30	40	50	60
	orf11ng-1.pep	DTVQAVIDEKSGDLRLRLTLKLYKATGDENKPFVLFQDGKEYTYVAQSELLDAQGNNILKG					
	orf11-1	DTVQAVIDEKSGDLRLRLTLKLYKATGDENKPFILFGDGKEYTYVAQSELLDAQGNNILKG					
15		70	80	90	100	110	120
	orf11ng-1.pep	DTVQAVIDEKSGDLRLRLTLKLYKATGDENKPFVLFQDGKEYTYVAQSELLDAQGNNILKG					
	orf11-1	DTVQAVIDEKSGDLRLRLTLKLYKATGDENKPFILFGDGKEYTYVAQSELLDAQGNNILKG					
		70	80	90	100	110	120
	orf11ng-1.pep	IGFSAPKKQYTLNGDTVEVRLSAPETNGLKIDKVTFTKDSYLVNVRFDIANGSGQTANL					
	orf11-1	IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVTFTKGSYLVNVRFDIANGSGQTANL					
20		130	140	150	160	170	180
	orf11ng-1.pep	IGFSAPKKQYTLNGDTVEVRLSAPETNGLKIDKVTFTKDSYLVNVRFDIANGSGQTANL					
	orf11-1	IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVTFTKGSYLVNVRFDIANGSGQTANL					
		130	140	150	160	170	180
	orf11ng-1.pep	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAKSGKSEAEYIRKT					
	orf11-1	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAKSGKSEAEYIRKT					
25		190	200	210	220	230	240
	orf11ng-1.pep	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAKSGKSEAEYIRKT					
	orf11-1	SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDAKSGKSEAEYIRKT					
		190	200	210	220	230	240
	orf11ng-1.pep	PTGWLGMIEHHFMSTWILQPKGGQNVCAQGDICRIDIKRRNDKLYSASVSPLTAIPTRG					
	orf11-1	PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSPLAAIQN-GA					
30		250	260	270	280	290	300
	orf11ng-1.pep	PTGWLGMIEHHFMSTWILQPKGGQNVCAQGDICRIDIKRRNDKLYSASVSPLTAIPTRG					
	orf11-1	PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSPLAAIQN-GA					
		250	260	270	280	290	
	orf11ng-1.pep	KPKMAVNLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIV					
	orf11-1	KAEASINLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAII					
40		310	320	330	340	350	360
	orf11ng-1.pep	KPKMAVNLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIV					
	orf11-1	KAEASINLYAGPQTTSVIANIADNLQAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAII					
		310	320	330	340	350	
	orf11ng-1.pep	VLTIIIVKAVLYPLTNASYRSMAKMRAAPKLQTIKEYGDDRMAQQQAMMQLYKDEKINP					
	orf11-1	VLTIIIVKAVLYPLTNASYRSMAKMRAAPKLQAIKEYGDDRMAQQQAMMQLYTDEKINP					
45		360	370	380	390	400	410
	orf11ng-1.pep	VLTIIIVKAVLYPLTNASYRSMAKMRAAPKLQTIKEYGDDRMAQQQAMMQLYKDEKINP					
	orf11-1	VLTIIIVKAVLYPLTNASYRSMAKMRAAPKLQAIKEYGDDRMAQQQAMMQLYTDEKINP					
		360	370	380	390	400	410
	orf11ng-1.pep	LGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPIIMAATMFAQT					
	orf11-1	LGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPIIMAATMFAQT					
50		420	430	440	450	460	470
	orf11ng-1.pep	LGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPIIMAATMFAQT					
	orf11-1	LGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYIILPIIMAATMFAQT					
		420	430	440	450	460	470
	orf11ng-1.pep	YLNPPPTDPMQAKMMKIMPLVFSVMFFFEFAGLVLYWVNNLLTIAQQWHINRSIEKQRA					
	orf11-1	YLNPPPTDPMQAKMMKIMPLVFSVMFFFEFAGLVLYWVNNLLTIAQQWHINRSIEKQRA					
55		480	490	500	510	520	530
	orf11ng-1.pep	YLNPPPTDPMQAKMMKIMPLVFSVMFFFEFAGLVLYWVNNLLTIAQQWHINRSIEKQRA					
	orf11-1	YLNPPPTDPMQAKMMKIMPLVFSVMFFFEFAGLVLYWVNNLLTIAQQWHINRSIEKQRA					
		480	490	500	510	520	530
	orf11ng-1.pep	QGEVVSX					
	orf11-1	QGEVVSX					
60		540					

65 In addition, ORF11ng-1 shows significant homology with an inner-membrane protein from the database (accession number p25754):

10

15

20

25

30

35

40

45

50

55

60

65

Based on this analysis, including the homology to an inner-membrane protein from *P. putida* and the predicted transmembrane domains (seen in both the meningococcal and gonococcal proteins), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 8

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 59>:

```

1  ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
51  NAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101 CGCCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGCGCT GGGTATTNNG
151 TTCGTACACG CCAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
201 GGATTGGGAT GCCGGACAAT ATGTCGAAAT CCTCCGNCAC ACAGGCGGCA
251 ACCGTTACGA AGTT.TTTAT CGCGGTACG. ACTGGCAGGC TCAAAATACG
301 GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA
351 AGGCAACCTT CTTATTATCA CACACCCTTA A

```

15 This corresponds to the amino acid sequence <SEQ ID 60; ORF13>:

```

1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
51  FVHAKTAVRK VETDSYQDLG AGQYVEILRH TGGNRYEVXY RGTWQAQNT
101 GQEELEPGTR ALIVRKEGNL LIITHP*

```

Further sequence analysis elaborated the DNA sequence slightly <SEQ ID 61>:

```

20 1  ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
51  nAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101 CGCCTGCCGC CGTCTTGACC GnCGCTCTGC TTTCCGCGCT GGGTATTNNG
151 TTCGTACACG CCAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
201 GGATTGGGAT GCCGGACAAT ATGTCGAAAT CCTCCGACAC ACAGGCGGCA
25 251 ACCGTTACGA AGTTTTtTAT CGCGGTACGc ACTGGCAGGC TCAAAATACG
301 GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA
351 AGGCAACCTT CTTATTATCA CACACCCTTA A

```

This corresponds to the amino acid sequence <SEQ ID 62; ORF13-1>:

```

30 1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
51  FVHAKTAVRK VETDSYQDLG AGQYVEILRH TGGNRYEVFY RGTWQAQNT
101 GQEELEPGTR ALIVRKEGNL LIITHP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF13 shows 92.9% identity over a 126aa overlap with an ORF (ORF13a) from strain A of *N.*

35 *meningitidis*:

```

                                     10      20      30      40      50
or13.pep      AVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT XALLSALGIXF
               |||||||
40 or13a      MTVWFVA AVAVLIIELLTGTVYLLVVSAA LAGSGIAYGLTGSTPAAVLT AALLSALGIWF
               10      20      30      40      50      60
               60      70      80      90      100     110
or13.pep      VHAKTAVRK VETDSYQDLG AGQYVEILRH TGGNRYEVXYRGTWQAQNTGQEELEPGTRA
               |||||||
45 or13a      VHAKTAVGK VETDSYQDLG AGYAEILRHAGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
               70      80      90      100     110     120
               120
or13.pep      LIVRKEGNLLIITHPX
50            |||||||::||

```

orf13a LIVRKEGNLLIIAKPX
130

The complete length ORF13a nucleotide sequence <SEQ ID 63> is:

```

5      1  ATGACTGTAT  GGTGTGTGTC  CGCTGTTGCC  GTCTTAATCA  TCGAATTATT
      51  GACGGGAACG  GTTTATCTTT  TGGTTGTCAG  CGCGGCTTTG  GCGGGTTCGG
     101  GCATTGCTTA  CGGGCTGACC  GGCAGCACGC  CTGCCGCCGT  CTTGACCGCC
     151  GCTCTGCTTT  CCGCGCTGGG  TATTTGGTTC  GTACACGCCA  AAACCGCCGT
     201  GGGAAAAGTT  GAAACGGATT  CATATCAGGA  TTTGGATGCC  GGGCAATATG
     251  CCGAAATCCT  CCGGCACGCA  GCGGGCAACC  GTTACGAAGT  TTTTATCGC
    301  GGTACGCACT  GGCAGGCTCA  AAATACGGGG  CAAGAAGAGC  TTGAACCAGG
    351  AACGCGCGCC  CTAATCGTCC  GCAAGGAAGG  CAACCTTCTT  ATCATCGCAA
    401  AACCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 64>:

```

15      1  MTVWFVAAVA  VLIIELLTGT  VYLLVVSAL  AGSGIAYGLT  GSTPAAVLTA
      51  ALLSALGIWF  VHAKTAVGKV  ETDSYQDLDA  GQYAEILRHA  GGNRYEVFGR
     101  GTHWQAQNTG  QEELEPGTRA  LIVRKEGNLL  IIAKP*

```

ORF13a and ORF13-1 show 94.4% identity in 126 aa overlap

```

20      orf13a.pep      10      20      30      40      50      60
      orf13-1          MTVWFVAAVAVLIIELLTGTVYLLVVSALAGSGIAYGLTGSTPAAVLTAALLSALGIWF
      orf13-1          |||||
                        10      20      30      40      50
      orf13a.pep      70      80      90      100     110     120
      orf13-1          VHAKTAVGKVETDSYQDLDAQYAEILRHAGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
      orf13-1          |||||
                        60      70      80      90      100     110
      orf13a.pep      130
      orf13-1          LIVRKEGNLLIIAKPX
      orf13-1          |||||
                        120
      orf13-1          LIVRKEGNLLIITHPX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF13 shows 89.7% identity over a 126aa overlap with a predicted ORF (ORF13.ng) from *N.gonorrhoeae*:

```

40      orf13          AVLIIELLTGTVYLLVVSALAGSGIAYGLTGSTPAAVLTXALLSALGIXF  51
      orf13ng          |||||
      orf13ng          MTVWFVAAVAVLIIELLTGTVYLLVVSALAGSGIAYGLTGSTPAAVLTAALLSALGIWF  60
      orf13          VHAKTAVRKVETDSYQDLDAQYVEILRHTGGNRYEVXYRGTXWQAQNTGQEELEPGTRA  111
      orf13ng          |||||
      orf13ng          VHAKTAVGKVETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA  120
      orf13          LIVRKEGNLLIITHP  126
      orf13ng          |||||
      orf13ng          LIVRKEGNLLIANP  135

```

50 The complete length ORF13ng nucleotide sequence <SEQ ID 65> is:

```

      1  ATGACTGTAT  GGTGTGTGTC  CGCTGTTGCC  GTCTTAATCA  TCGAATTATT
     51  GACGGGAACG  GTTTATCTTT  TGGTTGTCAG  CGCGGCTTTG  GCGGGTTCGG
    101  GCATTGCCTA  CGGGCTGACT  GGCAGCACGC  CTGCCGCCGT  CTTGACCGCC
    151  GCACTGCTTT  CCGCGCTGGG  CATTTGGTTC  GTACATGCCA  AAACCGCCGT
    201  GGGAAAAGTT  GAAACGGATT  CATATCAGGA  TTTGGATACC  GGAAAATATG
    251  CCGAAATCCT  CCGATACACA  GCGGGCAACC  GTTACGAAGT  TTTTATCGC
    301  GGTACGCACT  GGCAGGCGCA  AAATACGGGG  CAGGAAGTGT  TTGAACCGGG
    351  AACGCGCGCC  CTCATCGTCC  GCAAAGAAGG  TAACCTTCTT  ATCATCGCAA
    401  ACCCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 66>:

```

1  MTVWFVAAVA VLIIELLTGT VYLLVVSAA AGSGIAYGLT GSTPAAVLTA
51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDT GKAEILRYT GGNRYEVFYR
101 GTHWQAQNTG QEVFEPGTRA LIVRKEGNLL IIANP*

```

5 ORF13ng shows 91.3% identity in 126 aa overlap with ORF13-1:

```

10 orf13-1.pep      10      20      30      40      50
    AVLIIELLTGTVYLLVVSAAAGSGIAYGLTGSTPAAVLTXALLSALGIXF
    |||||
orf13ng      10      20      30      40      50      60
    MTVWFVAAVAVLIIELLTGTVYLLVVSAAAGSGIAYGLTGSTPAAVLTAALLSALGIWF

15 orf13-1.pep      60      70      80      90      100      110
    VHAKTAVRKVETDSYQDLDTAGQYVEILRHTGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
    |||||
orf13ng      60      70      80      90      100      110      120
    VHAKTAVGKVETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA

20 orf13-1.pep      120
    LIVRKEGNLLIITHPX
    |||||
orf13ng      120
    LIVRKEGNLLIIANPX
    130

```

Based on this analysis, including the extensive leader sequence in this protein, it is predicted that
25 ORF13 and ORF13ng are likely to be outer membrane proteins. It is thus predicted that the proteins
from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines
or diagnostics, or for raising antibodies.

Example 9

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 67>:

```

30 1  ATGTWTGATT TCGGTTTGG CGArCTGGTT TTTGTCGGCA TTATCGCCCT
    51  GATWgtCCTC GGCCCCGAAC GCsTGCCCGA GGCCGCCCGC AyCGCCGGAC
    101 GGcTCATCGG CAGGCTGCAA CGCTTTGTCG GcAGCGTCAA ACAGGAATTT
    151 GACACTCAAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
    201 AGCTGCCGcC GCTCAGGTTT GAGACAGCCT CAAAGAAACC GGTACGGATA
35 251 TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
    301 CTGCCCCGAAC AGCGGACACC TGCCGATTTT GGTGTCGATG AAAACGGCAA
    351 TCCGCT.TCC CGATGCGGCA AACACCCTAT CAGACGGCAT TTCCGACGTT
    401 ATGCCGTC..

```

This corresponds to the amino acid sequence <SEQ ID 68; ORF2>:

```

40 1  MXDFGLGELV FVGIIALIVL GPERXPEAAR XAGRLIGRLQ RFVGSVKQEF
    51  DTQIELEELR KAKQEFEEAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
    101 LPEQRTPADF GVDENGNPKS RCGKHPIRRH FFRYAV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 69>:

```

45 1  ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCGGCA TTATCGCCCT
    51  GATTGTCCTC GGCCCCGAAC GCCTGCCCGA GGCCGCCCGC ACCGCCGGAC
    101 GGCTCATCGG CAGGCTGCAA CGCTTTGTCG GCAGCGTCAA ACAGGAATTT
    151 GACACTCAAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
    201 AGCTGCCGCC GCTCAGGTTT GAGACAGCCT CAAAGAAACC GGTACGGATA
50 251 TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
    301 CTGCCCCGAAC AGCGGACACC TGCCGATTTT GGTGTCGATG AAAACGGCAA
    351 TCCGCTTCCC GATGCGGCAA ACACCCTATC AGACGGCATT TCCGACGTTA
    401 TGCCGTCGCA ACGTTCCTAC GCTTCCGCCG AAACCCTTGG GGACAGCGGG
    451 CAAACCGGCA GTACAGCGGA ACCCGCGGAA ACCGACCAAG ACCGCGCATG
    501 GCGGGAATAC CTGACTGCTT CTGCGGCCGC ACCCGTCGTA CAGACCGTCG

```

551 AAGTCAGCTA TATCGATACT GCTGTTGAAA CGCCTGTTCC GCACACCACT
 601 TCCCTGCGCA AACAGGCAAT AAGCCGCAAA CGCGATTTTC GTCCGAAACA
 651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

This corresponds to the amino acid sequence <SEQ ID 70; ORF2-1>:

5 1 MFD FGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF
 51 DTQIELEELR KAKQEFEEAAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
 101 LPEQRTPADF GVDENGPNLP DAANTLSDGI SDVMPERSY ASAETLGDSG
 151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETVPVHTT
 201 SLRKQAI SRK RDLRPKSRK PKLRVRKS*

10 Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 71 >:

1 ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCGGCA TTATCGCCCT
 51 GATTGTCCTC GGCCCCGAAC GCCTGCCCGA GGCCGCCCGC ACCGCCGGAC
 101 GGCTCATCGG CAGGCTGCAA CGCTTTGTCT GCAGCGTCAA ACAGGAATTT
 151 GACACGCAAA TCGAACTGGA AGAACTAAGG AAGGCAAAGC AGGAATTTGA
 15 201 AGCTGCCGCT GCTCAGGTTT GAGACAGCCT CAAAGAAACC GGTACGGATA
 251 TGGAGGGTAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
 301 CTGCCCGAAC AGCGCACGCC TGCTGATTTT GGTGTCGATG AAAACGGCAA
 351 TCCCTTTCCC GATGCGGCAA ACACCTATT AGACGGCATT TCCGACGTTA
 401 TGCCGTCCGA ACGTTCTTAC GCTTCCGCGG AAACCTTGG GGACAGCGGG
 20 451 CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGTGCATG
 501 GCGGGAATAC CTGACTGCTT CTGCCGCGGC ACCCGTCGTA CAGACCGTCG
 551 AAGTCAGCTA TATCGATACC GCTGTTGAAA CCCCTGTTCC GCATACCACT
 601 TCGTGCGTA AACAGGCAAT AAGCCGCAAA CGCGATTTGC GTCCTAAATC
 651 CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA

25 This encodes a protein having amino acid sequence <SEQ ID 72; ORF2a>:

1 MFD FGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF
 51 DTQIELEELR KAKQEFEEAAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
 101 LPEQRTPADF GVDENGPNFP DAANTLLDGI SDVMPERSY ASAETLGDSG
 151 QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETVPVHTT
 30 201 SLRKQAI SRK RDLRPKSRK PKLRVRKS*

The originally-identified partial strain B sequence (ORF2) shows 97.5% identity over a 118aa overlap with ORF2a:

		10	20	30	40	50	60
35	orf2.pep	MXDFGLGELVFVGIIALIVLGPXRPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR					
	orf2a	MFD FGLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR					
		10	20	30	40	50	60
40	orf2.pep	KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNXS					
	orf2a	KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNFP					
		70	80	90	100	110	120
45	orf2.pep	RCGKHPIRRHFRRYAV					
	orf2a	DAANTLLDGISDVMPERSYASAETLGDSGQTGSTAEPATDQDRAWREYLTASAAAPVV					
		130	140	150	160	170	180

50 The complete strain B sequence (ORF2-1) and ORF2a show 98.2% identity in 228 aa overlap:

	orf2a.pep	MFD FGLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR	60
	orf2-1	MFD FGLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR	60
55	orf2a.pep	KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNFP	120
	orf2-1	KAKQEFEEAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGPNLP	120
60	orf2a.pep	DAANTLLDGISDVMPERSYASAETLGDSGQTGSTAEPATDQDRAWREYLTASAAAPVV	180

```

      orf2-1      DAANTLSDGISDVMPERSYASAETLGDSSGQTGSTAETDQDRAWREYLTASAAAPVV 180
      orf2a.pep   QTVEVSYIDTAVETVPVHTTSLRKQAISRKRDLRPKSRAKPKLRVRKXS 229
      5          orf2-1      QTVEVSYIDTAVETVPVHTTSLRKQAISRKRDLRPKHRAKPKLRVRKXS 229

```

Further work identified a partial DNA sequence <SEQ ID 73> in *N.gonorrhoeae* encoding the following amino acid sequence <SEQ ID 74; ORF2ng>:

```

      1  MFD FGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL
      51  DTQIELEELR KVKQAFEAAA AQVRDSLKET DTMQNSLHD ISDGLKPWEK
      101 LPEQRTPADF GVDEKGNLSL RYGKHIRRH FRRYAV-

```

Further work identified the complete gonococcal gene sequence <SEQ ID 75>:

```

      1  ATGTTTGATT TCGGTTTGGG CGAGCTGATT TTTGTGGGCA TTATCGCCCT
      51  GATTGTCCTT GGTCCAGAAC GCCTGCCCGA AGCCGCCCGC ACTGCCGGAC
      15  101  GGCTTATCGG CAGGCTGCAA CGCTTTGTAG GAAGCGTCAA ACAAGAACTT
      151  GACACTCAAA TCGAACTGGA AGAGCTGAGG AAGGTCAAGC AGGCATTCTGA
      201  AGCTGCCGCC GCTCAGGTTT GAGACAGCCT CAAAGAAACC GATACGGATA
      251  TGCAGAACAG TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
      301  CTGCCCGAAC AGCGCACGCC tgccgatttc gGTGTCGATg AAAacggcaa
      20  351  tcccccttccc gATACGGCAA ACACCGTATC AGACGGCATT TCCGACGTTA
      401  TGCCGTCTGA ACGTTCGGAT ACTtccgcCG AAACCTTGG GGACGACAGG
      451  CAAACCGGCA GTACAGCCGA ACCTGCGGAA ACCGACAAAG ACCGCGCATG
      501  GCGGGAATAC CTGactgctt ctgcccgcgc acctgtcgta Cagagggccg
      551  tcgaagtcag ctaTATCGAT ACTGCTGTG AAacgcctgT tccgcaCacc
      25  601  acttccctgc gcaAACAGGC AATAAACCGC AAACGCGATT TttgtccgaA
      651  ACACCGCGCc aAACCGAAat tgcgcgtcCG TAAATCATAA

```

This encodes a protein having the amino acid sequence <SEQ ID 76; ORF2ng-1>:

```

      1  MFD FGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL
      51  DTQIELEELR KVKQAFEAAA AQVRDSLKET DTMQNSLHD ISDGLKPWEK
      30  101  LPEQRTPADF GVDENG NPLP DTANTVSDGI SDVMPERSD TSAETLGDDR
      151  QTGSTAEPAE TDKDRAWREY LTASAAAPVV QRAVEVSYID TAVETVPVHT
      201  TSLRKQAINR KRDFCPKHRA KPKLRVRKS*

```

The originally-identified partial strain B sequence (ORF2) shows 87.5% identity over a 136aa overlap with ORF2ng:

```

      35  orf2.pep      MXDFGLGELVFVGIIALIVLGPXRPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR 60
      orf2ng      MFD FGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60

      40  orf2.pep      KAKQFEAAAAQVRDSLKETGDTMEGNLHDISDGLKPWEKLEQRTPADFGVDENG NPLS 120
      orf2ng      KVKQAFEAAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLEQRTPADFGVDEKGNLSL 120

      orf2.pep      RCGKHPIRRHFRRYAV 136
      45  orf2ng      RYGKHIRRHFRYAV 136

```

The complete strain B and gonococcal sequences (ORF2-1 & ORF2ng-1) show 91.7% identity in 229 aa overlap:

```

      10      20      30      40      50      60
      50  orf2-1.pep    MFD FGLGELVFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR
      orf2ng-1      MFD FGLGELIFVGIIALIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR
      10      20      30      40      50      60

      70      80      90      100     110     120
      55  orf2-1.pep    KAKQFEAAAAQVRDSLKETGDTMEGNLHDISDGLKPWEKLEQRTPADFGVDENG NPLP
      orf2ng-1      KVKQAFEAAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLEQRTPADFGVDENG NPLP

```

-97-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf2-1.pep	DAANTLS	SDG	ISDV	MP	SERS	YASAE
	orf2ng-1	DTANTV	SDG	ISDV	MP	SERS	YASAE
		130	140	150	160	170	180
10	orf2-1.pep	Q-TVEV	SYID	TAVET	PVPHT	TS	LRKQ
	orf2ng-1	QRAVEV	SYID	TAVET	PVPHT	TS	LRKQ
		190	200	210	220	229	-

Computer analysis of these amino acid sequences indicates a transmembrane region (underlined),
 and also revealed homology (59% identity) between the gonococcal sequence and the TatB protein
 of *E.coli*:

gnl|PID|el292181 (AJ005830) TatB protein [Escherichia coli] Length = 171
 Score = 56.6 bits (134), Expect = 1e-07
 Identities = 30/88 (34%), Positives = 52/88 (59%), Gaps = 1/88 (1%)

20 Query: 1 MFD FGLGELIFVGIILIVLGPRLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60
 MFD G EL+ V II L+VLGP+RLP A +T I L+ +V+ EL +++L+E +
 Sbjct: 1 MFDIGFSELLLVFIIGLVVLPQRLPVAVKTVAGWIRALRSLATTVQNELTQELKLQEFQ 60

25 Query: 61 -KVKQAFEAAAAQVRDSLKETDTDMQNS 87
 +K+ +A+ + LK + +++ +
 Sbjct: 61 DSLKKVEKASLTNLTPELKASMDLRQA 88

Based on this analysis, it was predicted that ORF2, ORF2a and ORF2ng are likely to be membrane
 proteins and so the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be
 useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF2-1 (16kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above.
 The products of protein expression and purification were analyzed by SDS-PAGE. Figure 3A
 shows the results of affinity purification of the GST-fusion protein, and Figure 3B shows the results
 of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice,
 whose sera were used for Western blots (Figure 3C), ELISA (positive result), and FACS analysis
 (Figure 3D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is
 a useful immunogen.

Example 10

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 77>:

40 1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51 CGC.TGCGGG ACACGTGACAG GTATTCCATC GCATGGCGgA GKTAACgCT
 101 TTgCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTAAAA
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
 201 CACTATGGGC GACCAAGGTT CAGGcAGTTT GACAGGGGGG TCGCTACTCC
 251 ATTGATGCAC kGrTwCsTGG CGAATACATA AACAGCCCTG CCGTCCGTAC
 301 CGATTACACC TATCCACGTT ACGAAACCAC CGCTGAAACA ACATCAGGCG
 351 GTTTGACAGG TTAAACCACT TCTTTATCTA CACTTAATGC CCCTGCACTC
 401 TCTCGACCCC AATCAGACGG TAGCGGAAGT AAAAGCAGTC TGGGCTTAAA
 451 TATTGGCGGG ATGGGGGATT ATCGAAATGA AACCTTGACG ACTAACCCGC

501 GCGACACTGC CTTTCTTTCC CACTTGGTAC AGACCGTATT TTTCCTGCGC
 551 GGCATAGACG TTGTTTCTCC TGCCAATGCC GATACAGATG TGTTTATTAA
 601 CATCGACGTA TTCGGAACGA TACGCAACAG AACCGAAATG..

This corresponds to the amino acid sequence <SEQ ID 78; ORF15>:

5 1 MQARLLIPIL FSVFILLSACG TLTGIPSHGG XKRFAVEQEL VAASARAAVK
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDAXXG EYINSPAVRT
 101 DYTYPYRYETT AETTSGLTGG LTTSLSTLNA PALSRQSDG SGSKSSLGLN
 151 IGGMGDYRNE TLTTNPRDTA FLHLVQTVF FLRGIDVVSF ANADTDVFIN
 201 IDVFGTIRNR TEM..

10 Further work revealed the complete nucleotide sequence <SEQ ID 79>:

1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51 CGCCTGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GGTAAACGCT
 101 TTGCGGTGCA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTAAAA
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
 15 201 CACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
 251 TTGATGCACT GATTCTGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
 301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
 351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
 401 CTCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
 20 451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
 501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATT TTCTGCGCG
 551 GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAAC
 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
 651 TGCCGAAACA CTGAAAGCCC AAACAAAACCT GGAATATTTC GCAGTAGACA
 25 701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
 751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
 801 AGGAATTAAA CCGACGGAAG GATTAATGGT CGATTCTCC GATATCCGAC
 851 CATACGGCAA TCATACGGGT AACTCCGCC CATCCGTAGA GGCTGATAAC
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GTAGTGCAGC AACATAGACA
 30 951 AGGACAACCT TGA

This corresponds to the amino acid sequence <SEQ ID 80; ORF15-1>:

1 MQARLLIPIL FSVFILLSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK
 51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT
 101 DYTYPYRYETT AETTSGLTGG LTTSLSTLNA PALSRQSDG SGSKSSLGLN
 151 IGGMGDYRNE TLTTNPRDTA FLHLVQTVF FLRGIDVVSF ANADTDVFIN
 201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
 251 AYKENYALWM GPYKVSIGIK PTEGLMVDFS DIRPYGNHTG NSAPSVEADN
 301 SHEGYGYSDE VVRQHRQGP *

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 81>:

40 1 ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51 CGCCTGCGGG AACTGACAG GTATTCCATC GCATGGCGGA GGTAAACGCT
 101 TTGCGGTGCA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTAAAA
 151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
 45 201 AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
 251 TTGATGCACT GATTCTGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
 301 GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
 351 TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
 401 CGCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
 451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
 50 501 CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATT TTCTGCGCG
 551 GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACGGATGT GTTTATTAAAC
 601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
 651 TGCCGAAACA CTGAAAGCCC AAACAAAACCT GGAATATTTC GCAGTAGACA
 701 GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
 55 751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGACCGTATA AAGTAAGCAA
 801 AGGAATTAAA CCGACAGAAG GATTAATGGT CGATTCTCC GATATCCAAC
 851 CATACGGCAA TCATATGGGT AACTCTGCC CATCCGTAGA GGCTGATAAC
 901 AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCAGC GACATAGACA
 951 AGGGCAACCT TGA

60 This encodes a protein having amino acid sequence <SEQ ID 82; ORF15a>:

1 MQARLLIPIL FSVFILLSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK

51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT
 101 DYTYPRIYETT AETTSGGLTG LTSLSTLNA PALSRQSDG SGSKSSLGLN
 151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVS ANADTDVFIN
 201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
 251 AYKENYALWM GPYKVSQGIK PTEGLMVDFS DIQPYGNHMG NSAPSVEADN
 301 SHEGYGYSDE AVRRHRQGPQ *

The originally-identified partial strain B sequence (ORF15) shows 98.1% identity over a 213aa overlap with ORF15a:

10	orf15.pep	10 20 30 40 50 60	MQARLLIPILFSVFILSACGTLTGIPSHGGKRFQELVAASARAQVMDLQALHGR
	orf15a	10 20 30 40 50 60	MQARLLIPILFSVFILSACGTLTGIPSHGGKRFQELVAASARAQVMDLQALHGR
15	orf15.pep	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGRYSIDAXXXGEYINSPAVRTDYTYPRIYETTAETTSGGLTG
	orf15a	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRIYETTAETTSGGLTG
20	orf15.pep	130 140 150 160 170 180	LTSLSTLNAAPALSRQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
	orf15a	130 140 150 160 170 180	LTSLSTLNAAPALSRQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
25	orf15.pep	190 200 210	FLRGIDVVSANADTDVFINIDVFGTIRNRTEM
30	orf15a	190 200 210 220 230 240	FLRGIDVVSANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL

The complete strain B sequence (ORF15-1) and ORF15a show 98.8% identity in 320 aa overlap:

35	orf15a.pep	10 20 30 40 50 60	MQARLLIPILFSVFILSACGTLTGIPSHGGKRFQELVAASARAQVMDLQALHGR
	orf15-1	10 20 30 40 50 60	MQARLLIPILFSVFILSACGTLTGIPSHGGKRFQELVAASARAQVMDLQALHGR
40	orf15a.pep	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRIYETTAETTSGGLTG
	orf15-1	70 80 90 100 110 120	KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRIYETTAETTSGGLTG
45	orf15a.pep	130 140 150 160 170 180	LTSLSTLNAAPALSRQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
	orf15-1	130 140 150 160 170 180	LTSLSTLNAAPALSRQSDGSGSKSSLGLNIGMGDYRNETLTNPRDTAFLSHLVQTVF
50	orf15a.pep	190 200 210 220 230 240	FLRGIDVVSANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
	orf15-1	190 200 210 220 230 240	FLRGIDVVSANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
55	orf15a.pep	250 260 270 280 290 300	IKPKTNAFEAAAYKENYALWMGPYKVSQGIKPTGLMVDFSQIOPYGNHMGNSAPSVEADN
60	orf15-1	250 260 270 280 290 300	IKPKTNAFEAAAYKENYALWMGPYKVSQGIKPTGLMVDFSQIOPYGNHMGNSAPSVEADN
65	orf15a.pep	310 320	SHEGYGYSDEAVRRHRQGPX
	orf15-1		SHEGYGYSDEVVRQHRQGPX

310

320

Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 83>:

```

      1 ATGCGGGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
      51 CGCCTGCGGG ACACGTGACAG GTATTCCATC GCATGGCGGA GGCAAACGCT
      101 TCGCGGTCTGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
      151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
      201 AACTATGGGC GACCAAGGTT CAGGCAAGTT GACAGGGGGT CGTACTCCA
      251 TTGATGCACT GATTTCGCGGC GAATACATAA ACAGCCCTGC CGTCCGCACC
      301 GATTACACCT ATCCGCGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
      351 TTTGACGGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
      401 CGCGCACCCA ATCAGACGGT AGCGGAAGTA GGAGCAGTCT GGGCTTAAAT
      451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CCAACCGCG
      501 CGACACTGCC TTTCTTTCCC ACTTGGTGCA GACCGTATTT TTCCTGCGCG
      551 GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAAC
      601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
      651 TGCCGAAACA CTGAAAGCCC AAACAAACT GGAATATTTC GCAGTAGACA
      701 GAACCAATAA AAAATTGCTC ATCAAACCCA AAACCAATGC GTTTGAAGCT
      751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
      801 AGGAATCAA CCGACGGAAG GATTGATGGT CGATTCTCC GATATCCAAC
      851 CATACGGCAA TCATACGGGT AACTCCGCC CATCCGTAGA GGCTGATAAC
      901 AGTCATGAG GGTATGGATA CAGCGATGAA GCAGTGCAG AACATAGACA
      951 AGGGCAACCT TGA
  
```

This encodes a protein having amino acid sequence <SEQ ID 84; ORF15ng>:

```

      1 MRARLLIPIL FSVFILSACG TLTGIPSHGG GKRFVEQEL VAASARAANK
      51 DMDLQALHGR KVALYIATMG DQSGSLTGG RYSIDALIRG EYINSPAVRT
      101 DYTYPRIYET AETTSGLTGT LTSLSTLNA PALSRQSDG SGRSSSLGLN
      151 IGGMGDYRNE TLTTNPRDTA FLHLVQTVF FLRGIDVVS ANADTDVFIN
      201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
      251 AYKENYALWM GPYKVSQGIK PTEGLMVDFF DIQPYGNHTG NSAPSVEADN
      301 SHEGYGYSDE AVRQHRQGP *
  
```

The originally-identified partial strain B sequence (ORF15) shows 97.2% identity over a 213aa overlap with ORF15ng:

```

      orf15.pep  MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFVEQELVAASARAANKDMDLQALHGR  60
      151 1:|||||
      orf15ng    MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFVEQELVAASARAANKDMDLQALHGR  60
      151 1:|||||
      orf15.pep  KVALYIATMGDQSGSLTGGRYSIDAXXXGEYINSPAVRTDYTYPRIYETTAETTSGLTGT  120
      401 1:|||||
      orf15ng    KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRIYETTAETTSGLTGT  120
      401 1:|||||
      orf15.pep  LTSLSTLNAPALSRQSDGSGSKSSSLGNIGMGDYRNETLTNPRDTAFLHLVQTVF  180
      451 1:|||||
      orf15ng    LTSLSTLNAPALSRQSDGSGSRSSSLGNIGMGDYRNETLTNPRDTAFLHLVQTVF  180
      451 1:|||||
      orf15.pep  FLRGIDVVSANADTDVFINIDVFGTIRNRTEM  213
      451 1:|||||
      orf15ng    FLRGIDVVSANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL  240
  
```

The complete strain B sequence (ORF15-1) and ORF15ng show 98.8% identity in 320 aa overlap:

```

      10      20      30      40      50      60
      50 orf15-1.pep  MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFVEQELVAASARAANKDMDLQALHGR
      orf15ng      MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFVEQELVAASARAANKDMDLQALHGR
      10      20      30      40      50      60
      55 orf15-1.pep  KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRIYETTAETTSGLTGT
      orf15ng      KVALYIATMGDQSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRIYETTAETTSGLTGT
      70      80      90      100     110     120
      60 orf15-1.pep  LTSLSTLNAPALSRQSDGSGSKSSSLGNIGMGDYRNETLTNPRDTAFLHLVQTVF
      orf15ng      LTSLSTLNAPALSRQSDGSGSKSSSLGNIGMGDYRNETLTNPRDTAFLHLVQTVF
  
```

```

|||||:|||||
orf15ng  LTTSLSLTLNAPALSRTQSDGSGSRSSLGLNIGGMGDYRNETLTTNPRDTAFLSHLVQTVF
          130          140          150          160          170          180

5         190          200          210          220          230          240
orf15-1.pep  FLRGIDVVSANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
|||||:|||||
orf15ng  FLRGIDVVSANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
          190          200          210          220          230          240

10        250          260          270          280          290          300
orf15-1.pep  IKPKTNAFEAAAYKENYALWMGPYKVSKGIKPTEGLMVDFSDIRPYGNHTGNSAPSVEADN
|||||:|||||
orf15ng  IKPKTNAFEAAAYKENYALWMGPYKVSKGIKPTEGLMVDFSDIQPYGNHTGNSAPSVEADN
          250          260          270          280          290          300

15        310          320
orf15-1.pep  SHEGYGYSDEVVRQHRQGPX
|||||:|||||
20        orf15ng  SHEGYGYSDEAVRQHRQGPX
          310          320

```

indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF15-1 (31.7kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 4A shows the results of affinity purification of the GST-fusion protein, and Figure 4B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 4C) and ELISA (positive result). These experiments confirm that ORFX-1 is a surface-exposed protein, and that it is a useful immunogen.

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 85>:

35	1	..GG.CAGCACA	AAAAACAGGC	GTTGAACGG	AAAAACCGTA	TTTACGATGA
	51	TGCCGGGTAT	GATATTCGGC	GTATTCACGG	GCGCATTCTC	CGCAAAATAT
	101	ATCCCCCGCT	TCGGGCTTCA	AATTTTCTTC	ATCCTGTTTT	TAACCCCGCT
	151	CGCATTCAAA	ACACTGCATA	CCGACCCTCA	GACGGCATCC	CGCCCCGCTG
40	201	CCGGACTGCC	CrGACTGACT	GCGGTTTCCA	CACGTGTCGG	CACAATGTCG
	251	AGCTGGGTG	GCATAGGCGG	CGTTCACTT	TCCGTCCCCT	TCTTAAATCCA
	301	CTGGGGCTTC	CCCCCCCATA	AAGCCATCGG	CACATCATCC	GGCCTTGCCT
	351	GGCCGATTGC	ACTCTCCGCG	GCAATATCGT	ATCTGCTCAA	CGGCCGTGAAT
45	401	ATTGCAGGAT	TGCCCCGAAG	GTACTGGGC	TTCCTTTACC	TGCCCCGCGT
	451	CGCCCTCCTC	AGCGGGCCAA	CCATTGCGTT	TGCCCCGCTC	GGGTGTCAAA
	501	CCGCCACAAA	ACTTCTCTCT	GCCAAACTCA	AAAAAATC.TT	CGGCATTATG
	551	TTGCTTTTGA	TTGCCGGAAG	AATGCTGTAC	AACCTGCTTT	AA

1 ..GQHKKQAVNG KTVFTMMPGM IFGVFTGAFS AKYIPAFGLQ IFFILFLTAV
51 AFKTLHTDPQ TASRPLPLGP XLTAVSTLFG TMSSWVGIGG GSLSVPFLIH

101 CGFPAHKAIG TSSGLAWPIA LSGAISYLLN GLNIAGLPEG SLGFLYLPVAV
 151 AVLSAATIAF APLGVKTAHK LSSAKLKKSF GIMLLLIAGK MLYNLL*

Further work revealed the complete nucleotide sequence <SEQ ID 87>:

5 1 ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCCGTAG GCAGTGGCGC
 51 AGGTTTATT GCCGGCCTGT TCGGCGTAGG CGGCGGCACG CTGATTGTCC
 101 CTGTCGTTTT ATGGGTGCTT GATTTGCAGG GTTGGGCACA ACATCCTTAC
 151 GCGCAACACC TCGCCGTCGG CACATCCTTC GCCGTCATGG TCTTCACCGC
 201 CTTTTCCAGT ATGCTGGGGC AGCACAAAAA ACAGGCGGTC GACTGGAAAA
 251 CCGTATTAC GATGATGCCG GGTATGATAT TCGGCGTATT CACGGGCGCA
 10 301 CTCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT
 351 GTTTTTAACC GCCGTCGCAT TCAAAACACT GCATACCGAC CCTCAGACGG
 401 CATCCCGCCC GCTGCCCGGA CTGCCCGGAC TGACTGCGGT TTCCACACTG
 451 TTCGGCACAA TGTCGAGCTG GGTCCGCATA GGCGGCGGTT CACTTTCGCT
 501 CCCCTTCTTA ATCCACTGCG GCTTCCCGCG CCATAAGACC ATCGGCACAT
 15 551 CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG
 601 CTCAACGGCC TGAATATTGC AGGATTGCCC GAAGGGTCAC TGGGCTTCCT
 651 TTACCTGCCG GCCGTGCGCG TCCTCAGCGC GGCAACCATT GCCTTTGCCG
 701 CGCTCGGTGT CAAAACCGCC CACAACTTT CTTCTGCCAA ACTCAAAAAA
 751 TC TTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT
 20 801 GCTTTAA

This corresponds to the amino acid sequence <SEQ ID 88; ORF17-1>:

1 MWHWDIILIL LAVGSAAGFI AGLFGVGGT LIVPVVLWVL DLQGLAQHPY
 51 AQHLAVGTSE AVMVETAFFS MLGQHKQAV DWKTVFTMMP GMIFGVFTGA
 25 101 LSAKYIPAFG LOIFFILFLT AVAFKTLHTD PQASRPLPG LPGLTAVSTL
 151 FGTSSWVGI GGSLSVFEL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
 201 LNGLNIAGLP EGSGLFLYLP AVAVLSAATI AFAPLVKTA HKLSSAKLKK
 251 XFGIMLLLIA GKMLYNLL*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical *H. influenzae* transmembrane protein HI0902 (accession number P44070)

30 ORF17 and HI0902 proteins show 28% aa identity in 192 aa overlap:

ORF17 3 HKKQAVNGKTVFTMMPGMIFGVFT-GAFSAKYIPAFGLQIF--FILFLTAVAFKTLHTDP 59
 HK + + V + P ++ VF G F + +IF +++L ++ D
 HI0902 72 HKLGNIVWQAVRILAPVIMLSVFICGLFIGRLDREISAKIFACLVVYLATKMVLSIKKD- 130
 35 ORF17 60 QTASRPLPGLPXLTAVSTLFGTMSWVGIGGSLVFLIHCGFPAHKAIGTSSGLAWPI 119
 Q ++ L L + L G SS GIGGG VPFL G +AIG+S+ +
 HI0902 131 QVTTKSLTPLSSVIG-GILIGMASSAAGIGGGFIVPFLTARGINIKQAIGSSAFCGMLL 189
 40 ORF17 120 ALSGAISYLLNGLNIAGLPEGSLGFLYLPVAVLSAATIAFAPLVGXXXXXXXXXXXXXXX 179
 +SG S++++G +PE SLG++YLPVAV ++A + + LG
 HI0902 190 GISGMFSFIVSGWGNPLMPEYSLGYIYLPVAVLGITATSFFTSKLGASATAKLPVSTLKKG 249
 ORF17 180 FGIMLLLIAGKM 191
 F + L+++A M
 45 HI0902 250 FALFLIVVAINM 261

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF17 shows 96.9% identity over a 196aa overlap with an ORF (ORF17a) from strain A of *N. meningitidis*:

50 orf17.pep 10 20 30
 GQHKQAVNGKTVFTMMPGMIFGVFTGAFS
 orf17a QGLAQHPYAQHLAVGT¹SAVMVETAFFSSMLGQHKQAVDWKTVFTMMPGMVFGVFAGALS
 50 60 70 80 90 100
 55 orf17.pep 40 50 60 70 80 90
 AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPXLTAVSTLFGTMSWVGIGG
 orf17a AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSWVGIGG

-103-

		110	120	130	140	150	160
		100	110	120	130	140	150
5	orf17.pep	GSLSVFPLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPVAV					
	orf17a	GSLSVFPLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPVAV					
		170	180	190	200	210	220
10	orf17.pep	AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLLX					
	orf17a	AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLLX					
		230	240	250	260		

The complete length ORF17a nucleotide sequence <SEQ ID 89> is:

15	1	ATGTGGCATT	GGGACATTAT	CTAATCCTG	CTTGCCGTAG	GCAGTGGCGC
	51	AGGTTTATT	GCCGGCCTGT	TCGGCGTAGG	CGGCGGCACG	CTGATTGTCC
	101	CTGTCGTTTT	ATGGGTGCTT	GATTTCGAGG	GTTTGGCACA	ACATCCTTAC
	151	GCGCAACACC	TCGCCGTCGG	CACATCCTTC	GCCGTCATGG	TCTTCACCCG
	201	CTTTTCCAGT	ATGCTGGGGC	AGCACAAAAA	ACAGGCGGTC	GACTGGAAAA
20	251	CCGTATTTAC	GATGATGCCG	GGTATGGTAT	TCGGCGTATT	CGCTGGCGCA
	301	CTCTCCGCAA	AATATATCCC	AGCGTTCGGG	CTTCAAATT	TCTTCATCCT
	351	GTTTTTAACC	GCCGTCGCAT	TCAAAACACT	GCATACCGAC	CCTCAGACGG
	401	CATCCCGCCC	GCTGCCCGGA	CTGCCCGGAC	TGACTGCGGT	TTCCACACTG
	451	TTCCGCACAA	TGTCGAGCTG	GGTCGGCATA	GGCGGCGGTT	CACTTCCGT
25	501	CCCCTTCTTA	ATCCACTGCG	GCTTCCCCGC	CCATAAGCC	ATCGGCACAT
	551	CATCCGGCCT	TGCCTGGCCG	ATTGCACTCT	CCGGCGCAAT	ATCGTATCTG
	601	CTCAACGGCC	TGAATATTGC	AGGATTGCCC	GAAGGGTCAC	TGGGCTTCCT
	651	TTACCTGCCC	GCCGTCGCCG	TCCTCAGCGC	GGCAACCATT	GCCTTTGCC
	701	CGCTCGGTGT	CAAAACGCC	CACAACTTT	CTTCTGCCAA	ACTCAAAAAA
30	751	TCCTTCGGCA	TTATGTTGCT	TTTGATTGCC	GGAAAAATGC	TGTACAACCT
	801	GCTTTAA				

This encodes a protein having amino acid sequence <SEQ ID 90>:

	1	MWHWDIILIL	LAVGSAAGFI	AGLFGVGGGT	LIVPVVLWVL	DLQGLAQHPY
	51	AQHLAGVTSF	AVMVFTAFSS	MLGQHKQAV	DWKTFTMMP	GMVFGVFAGA
35	101	LSAKYIPAFG	LQIFFILFLT	AVAFKTLHTD	PQTASRPLPG	LPGLTAVSTL
	151	FGTMSSWVGI	GGGSLVPFL	IHCGFPAHKA	IGTSSGLAWP	IALSGAISYL
	201	LNLGLNIAGLP	EGSLGFLYLP	AVAVLSAATI	AFAPLGVKTA	HKLSSAKLKK
	251	SFGIMLLLIA	GKMLYNLL*			

ORF17a and ORF17-1 show 98.9% identity in 268 aa overlap:

40	orf17a.pep	10	20	30	40	50	60
	orf17-1	10	20	30	40	50	60
45	orf17a.pep	70	80	90	100	110	120
	orf17-1	70	80	90	100	110	120
50	orf17a.pep	130	140	150	160	170	180
	orf17-1	130	140	150	160	170	180
55	orf17a.pep	190	200	210	220	230	240
	orf17-1	190	200	210	220	230	240
60	orf17a.pep	250	260	269			
	orf17-1	250	260	269			
65	orf17a.pep	HKLSSAKLKKSFGIMLLLIAGKMLYNLLX					

-104-

```

      |||||
orf17-1  HKLSSAKLKKXFGIMLLLIAGKMLYNLLX
          250      260

```

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF17 shows 93.9% identity over a 196aa overlap with a predicted ORF (ORF17.ng) from *N. gonorrhoeae*:

```

      orf17.pep                                GQHKQAVNGKTVFTMMPGMIFGVFTGAFS  30
      |||||: ||:|:|||||:|:|:|
10  orf17ng  QGLAQHPYAQHLAVGTSFAVMVFTAFSSMLGQHKQAVDWKTI FAMMPGMIFGVFAGALS  102
      orf17.pep  AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPXLTA VSTLFGTMSSWVGIGG  90
      |||||: ||:|:|||||:|:|:|
15  orf17ng  AKYIPAFGLQIFFILFLTAVAFKTLHTGRQTASRPLPGLPGLTA VSTLFGAMSSWVGIGG  162
      orf17.pep  GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPV  150
      |||||: ||:|:|||||:|:|:|
      orf17ng  GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLPV  202
20  orf17.pep  AVLSAATIAFAPLGVKTAHKLSSAKLKKSFSGIMLLLIAGKMLYNLL  196
      |||||: ||:|:|||||:|:|:|
      orf17ng  AVLSAATIAFAPLGVKTAHKLSSAKLKEFSGIMLLLIAGKMLYNLL  268

```

An ORF17ng nucleotide sequence <SEQ ID 91> is predicted to encode a protein having amino acid sequence <SEQ ID 92>:

```

25      1  MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
      51  AQHLAVGTSF AVMVFTAFSS MLGQHKQAV DWKTI FAMMP GMIFGVFAGA
      101  LSAKYIPAFG LQIFFILFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL
      151  FGAMSSWVG I GGSLSVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
      201  VNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE
30  251  SFGIMLLLIA GKMLYNLL*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 93>:

```

      1  ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCcgtag gcAGTGC GGC
      51  AGGTTTTATT GCCGGCCTGT Tcgggtgtagg cggcgGTACG CTGATTGTCC
101  CTGTCGTTTT ATGGGTGCTT GATTTCGAGG GTTTGGCACA ACATCCTTAC
35  151  GCGCAACACC TCGCCGTCGG CacaTccttc gcCGTCATGG TCTTCACCGC
      201  CTTTTCCAGT ATGTTGGGGC AGCACAAAAA ACAGGCGGTC GACTGGAAAA
      251  CCATATTTGC GATGATGCCG GGTATGATAT TCGGCGTATT CGCTGGCGCA
      301  CTCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT
      351  GTTTTAAACC GCCGTCGCAT TCAAAACACT GCATACCGGT CGTCAGACGG
40  401  CATCCCGCCC GCTGCCCGGG CTGCCCGGAC TGACTGCGGT TTCCACACTG
      451  TTCGGCGCAA TGTGAGCTG GGTCCGCATA GGCGGCGGTT CACTTTCGCT
      501  CCCCTTCTTA ATCCACTGCG GCTTCCCGCG CCATAAAGCC ATCGGCACAT
      551  CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG
      601  GTCAACGGTC TGAATATTGC AGGATTGCCG GAAGGGTCGC TGGGCTTCCT
45  651  TTACCTGCCC GCCGTCGCCG TCCTCAGCGC GGCAACCATT GCCTTTGCCG
      701  CGCTCGGTGT CAAAACCGCC CACAACTTT CTTCTGCCAA ACTCAAAGAA
      751  TCCTTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT
      801  GCTTTAA

```

This corresponds to the amino acid sequence <SEQ ID 94; ORF17ng-1>:

```

50      1  MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
      51  AQHLAVGTSF AVMVFTAFSS MLGQHKQAV DWKTI FAMMP GMIFGVFAGA
      101  LSAKYIPAFG LQIFFILFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL
      151  FGAMSSWVG I GGSLSVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
      201  VNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE
55  251  SFGIMLLLIA GKMLYNLL*

```

ORF17ng-1 and ORF17-1 show 96.6% identity in 268 aa overlap:

```

      10      20      30      40      50      60
orf17-1.pep  MWHWDIILILLAVGSAAGFIAGLFGVGGGT LIVPVVLWVL DLQGLAQHPYAQHLAVGTSF

```

	orfl7ng-1	 MWHWDIILILLAVGSAAGFIAGLFGVGGGTLIVPVVLWVLDLQGLAQHPYAQHLAVGTSF
		10 20 30 40 50 60
5	orfl7-1.pep	70 80 90 100 110 120 AVMVFTAFSSMLGQHKQAVDWKTIVFTMMPGMIFGVFTGALS AKYIPAFGLQIFFILFLT
	orfl7ng-1	 AVMVFTAFSSMLGQHKQAVDWKTIFAMMPGMIFGVFAGALS AKYIPAFGLQIFFILFLT
10		70 80 90 100 110 120
	orfl7-1.pep	130 140 150 160 170 180 AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGGGSLSVPFLIHCGFPAHKA
	orfl7ng-1	 AVAFKTLHTGRQTASRPLPGLPGLTAVSTLFGAMSSWVGIGGGSLSVPFLIHCGFPAHKA
15		130 140 150 160 170 180
	orfl7-1.pep	190 200 210 220 230 240 IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLP PAVAVLSAATIAFAPLGVKTA
20	orfl7ng-1	 IGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLP PAVAVLSAATIAFAPLGVKTA
		190 200 210 220 230 240
	orfl7-1.pep	250 260 269 HKLSSAKLKKXFGIMLLIAGKMLYNLLX
25	orfl7ng-1	 HKLSSAKLKESFGIMLLIAGKMLYNLLX
		250 260

In addition, ORF17ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

30 sp|P44070|Y902_HAEIN HYPOTHETICAL PROTEIN HI0902 pir||G64015 hypothetical protein
HI0902 - Haemophilus influenzae (strain Rd KW20) gi|1573922 (U32772) H. influenzae
predicted coding region HI0902 [Haemophilus influenzae] Length = 264
Score = 74 (34.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23
Identities = 15/43 (34%), Positives = 23/43 (53%)

35 Query: 55 AVGTSFAVMVFTAFSSMLGQHKQAVDWKTIFAMMPGMIFGVF 97
A+GTSFA +V T S HK + W+ + + P ++ VF
Sbjct: 52 ALGTSFATIVITGIGSAQRHHKLGNIWQAVRILAPVIMLSVF 94
Score = 195 (91.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23
Identities = 44/114 (38%), Positives = 65/114 (57%)

40 Query: 150 LFGAMSSWVGIGGGSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGL 209
L G SS GIGGG VPFL G +AIG+S+ + +SG S++V+G +
Sbjct: 148 LIGMASSAAGIGGGFIVPFLTARGINIKQAIGSSAFCGMLLGISGMFSFIVSGWGNPLM 207
45 Query: 210 PEGSLGFLYLP PAVAVLSAATIAFAPLGVKTAHKLSSAKLKESFGIMLLIAGKM 263
PE SLG++YLPVAV ++A + + LG KL + LK+ F + L+++A M
Sbjct: 208 PEYSLGYIYLPVAVLGITATSFFTSKLGASATAKLPVSTLKKGFALFLIVVAINM 261

50

This analysis, including the homology with the hypothetical *H. influenzae* transmembrane protein, suggests that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 12

55 The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 95>:

1 ..GGAAACGGAT GGCAGGCAGA CCCC GAACAT CCGCTGCTCG GGCTTTTTCG
51 CGTCAGTAAT GTATCGATGA CGCTTGCTTT TGTCCGAATA TGTGCGTTGG
101 TGCATTATTG CTTTTCGGGA ACGGTTCAAG TGTTTGTGTT TCGCGGCACTG
151 CTCAAACCTT ATGCGCTGAA GCCGGTTTAT TGGTTCGTGT TGCAGTTTGT
201 GCTGATGGCG GTTGCCTATG TCCACCGCTG CGGTATAGAC CGGCAGCCGC
60 251 CGTCAACGTT CGGCGGCTCG CAGCTGCGAC TCGGCGGGTT GACGGCAGCG

5 1 ..GNGWQADPEH PLLGLFAVSN VSMTLAFVGI CALVHYCFSG TVQVVFVAAL
51 LKLYALKPVY WFLVQFVLMA VAYVHRCGID RQPPSTFGGS QLRGLGLTAA
101 LMQVSVLVLL LSEIGR*

	1	ATGATTTTGC	TGCATTGGA	TTTTTGTCT	GCCTTACTGT	ATGCGGCGGT
10	51	TTTTCTGTTT	CTGATATTCC	GCGCAGGAAT	GTTGCAATGG	TTTTGGGCGA
	101	GTATTATGCT	GTGGCTGGGC	ATATCGGTTT	TGGGGGCAAA	GCTGATGCC
	151	GGCATATGGG	GAATGACCCG	CGCCGCGCCC	TGTGTCATCC	CCCATTTTTA
	201	CCTGACTTTG	GCAGCATAT	TTTTTTTCAT	CGGGCATTGG	AACCGGAAAA
	251	CAGATGAAAA	CGGATGGCAG	GCAGACCCCG	AACATCCGCT	GCTCGGGCTT
	301	TTTGCCGTC	GTAATGTATC	GATGACGCTT	TCTTTTGTCT	GAATATGTGC
15	351	GTTGGTGCAT	TATTGCTTTT	CGGGAACGGT	GCAAGTGTTT	GTTGTTGCGG
	401	CAC TGCTCAA	ACTTTATGCG	CTGAAGCCGG	TTTATTGGTT	CGTGTTCAG
	451	TTTGTGCTGA	TGGCGGTTGC	CTATGTCCAC	CGCTCGGGTA	TAGACGGGCA
	501	GCCGCCGTGA	ACGTTCCGGC	GCTCGCAGCT	GCGACTCGGC	GGGTTGACGG
	551	CAGCGTTGAT	GCAGGTCTCG	GTACTGGTGC	TGCTGCTTTC	AGAAATTGGA
20	601	AGATAA				

25

1	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIMLWLW	ISVLGAKLMP
51	GIWGMTRAAP	LFIPHFYLT	GSIFFFIGHW	NRKTDGNGWQ	ADPEHPLLGL
101	FAVSNVSMTL	AFVGICALVH	YCFSGTVQVF	VFAALLKLYA	LKPVYVFWLQ
151	FVLMAVAYVH	RCGIDRQPPS	TFGGSQRLRG	GLTAALMQVS	VLVLLLSEIG
201	R*				

30 *meningitidis:*

```

35      orf18.pep      GNGWQADPEHP10LLGLFAVSNV20SMTLAFVGI30
                        |||||
      orf18a      TRAAPLFIPHPFYLT60LGSIFFFI70GHWN80RKTDGNGWQADPEHP90LLGLFAVSNV100SMTLAFVGI110

40      orf18.pep      40CALVHYCFSGT50VQVFVFAALLKLYALKPVYWFV60LQFVLM70AVAYVHRCGIDRQ80PFSTFGGS90
                        |||||
      orf18a      CALVHYCF120SXTVQVFVFAALLKLYALKPVYWFV130LQFVLM140AVAYVHRCGIDRQ150PFSTFGGS160

45      orf18.pep      100QLRLGGLTAALMQVSVLV110LLSEIGRX
                        |||||
      orf18a      QLRLGGLTAALMQVSVLV180LLSEIGRX190

```

50	1	ATGATTTTGC	TGCATTTGGA	TTTTTTGTCT	GCCTTACTGT	ATGCGGCGGT
	51	TTTTCTGTTT	CTGATATTCC	GCGCAGGAAT	GTTGCAATGG	TTTTGGGCGA
	101	GTATTATGCT	GTGGCTGGGC	ATATCGGTTT	TGGGGGCAAA	GCTGATGCCA
	151	GGCATATGGG	GAATGACCCG	CGCCGCGCCC	TGTTTCATCC	CCCATTTTTA
	201	CCTGACTTTG	GGCAGCATAT	TTTTTTTCAT	CGGGCATTGG	AACCGGAAAA
55	251	CGGATGGAAA	CGGATGGCAG	GCAGACCCCG	AACATCTCTT	GCTCGGGCTG
	301	TTTGCCGTC	GTAATGTATC	GATGACGCTT	GCTTTTGTCT	GAATATGTGC
	351	GTTGGTGCAT	TATTGCTTTT	CGNGAACGGT	TCAAGTGTTT	GTGTTTGGCG
	401	CAC TGCTCAA	ACTTTATGCG	CTGAAGCCCG	TTTATTGGTT	CGTGTTCGAG

```

451 TTTGTGCTGA TGGCGGTTGC CTATGTCCAC CGCTGCGGTA TAGACCGGCA
501 GCCGCCGTCA ACGTTCGGCG GNTCGCAGCT GCGACTCGGC GGGTTGACGG
551 CAGCGTTGAT GCAGNTCTCG GTACTGGTGC TGCTGCTTTC AGAAATTGGA
601 AGATAA

```

5 This encodes a protein having amino acid sequence <SEQ ID 100>:

```

1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIMLWLG ISVLGAKIMP
51 GIWGMTRAAP LFIPHFYLTG LSIFFFIGHW NRKTDGNGWQ ADPEHPLLGL
101 FAVSNVSMTL AFVGICALVH YCFSXTVQVF VFAALLKLYA LKPVYWFVLQ
151 FVLMAYAVVH RCGIDRQPPS TFGGSQLRLG GLTAALMQXS VLVLLSEIG
201 R*

```

ORF18a and ORF18-1 show 99.0% identity in 201 aa overlap:

```

10 20 30 40 50 60
orfl8a.pep MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP
15 orfl8-1 MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP
10 20 30 40 50 60
20 70 80 90 100 110 120
orfl8a.pep LFIPHFYLTGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH
20 orfl8-1 LFIPHFYLTGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH
70 80 90 100 110 120
25 130 140 150 160 170 180
orfl8a.pep YCFSXTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGSQLRLG
orfl8-1 YCFSXTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGSQLRLG
130 140 150 160 170 180
30 190 200
orfl8a.pep GLTAALMQXSVLVLLSEIGRX
orfl8-1 GLTAALMQXSVLVLLSEIGRX
190 200
35

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF18 shows 93.1% identity over a 116aa overlap with a predicted ORF (ORF18.ng) from *N. gonorrhoeae*:

```

40 orfl8.pep GNGWQADPEHPLLGLFAVSNVSMTLAFVGI 30
orfl8ng TRAAPLFIPHFYLTGSIFFFIGYWNKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGI 115
45 orfl8.pep CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGS 90
orfl8ng CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAYAVVHRCGIDRQPPSTFGGS 175
orfl8.pep QLRLGGLTAALMQXSVLVLLSEIGR 116
orfl8ng QLRLGVLAAMLQVAVTAMLLAEIGR 201

```

50 The complete length ORF18ng nucleotide sequence is <SEQ ID 101>:

```

1 ATGATTTTGC TGCATTGGA TTTTGTCT GCCTTACTGt aTGCgGcggt
51 tttTctgTTT CTGATATTC GCGCAGGAAT GTTGCAATGG TTTTGGGCGA
101 GTATTGCGTT GTGGCTCGGC ATCTCGGTTT TAGGGGTAAa GCTGATGCCG
151 GGGATGTGGG GAATGACCCG CGCCGCGCCT TTGTTTCATCC CCCATTTTAA
55 201 CCTGACTTGG GGCAGCATAT TTTTTCAT CCGGTATTGG AACCGGAAAA
251 CAGATGGAAA CGGATGGCAG GCAGACCCCG AACATCCGCT GCTCGGCCTT
301 TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTTGTCG GAATATGTGC
351 GTTGGTGCAT TATTGCTTTT CGGGAACGGT TCAAGTGTTC GTGTTTGCAG
401 CATTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGGTT CGTGTTCAG
60 451 TTTGTATTGA TGGCGGtgc CTATGTCCAC CGCTGCGGTA TAGACCGGCA
501 GCCGCCGTCA ACGTTCGGCG GTTCGCAGCT GCGACTCGGC GTGTTGGCGG

```

-108-

551 CGATGTTGAT GCAGGTTGCG GTAACGGCGA TGCTGCTTGC CGAAATCGGC
601 AGATGA

This encodes a protein having amino acid sequence <SEQ ID 102>:

5 1 MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIALWLG ISVLGVKLMP
 51 GMWGMTRAAP LFIPHFYLT L GSIFFFIGYW NRKTDGNGWQ ADPEHPLLGL
 101 FAVSNVSMTL AFVGICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLO
 151 FVLMAYAYVH RCGIDRQPPS TFGGSQRLRG VLAAMLMOVA VTAMLLAEIG
 201 R*

This ORF18ng protein sequence shows 94.0% identity in 201 aa overlap with ORF18-1:

10		10	20	30	40	50	60
	orf18-1.pep	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIALWLG	ISVLGVKLMP	GMWGMTRAAP
	orf18ng	MILLHLDFLS	ALLYAAVFLF	LIFRAGMLQW	FWASIALWLG	ISVLGVKLMP	GMWGMTRAAP
15		70	80	90	100	110	120
	orf18-1.pep	LFIPHFYLT	L GSIFFFIGYW	NRKTDGNGWQ	ADPEHPLLGL	FAVSNVSMTL	AFVGICALVH
	orf18ng	LFIPHFYLT	L GSIFFFIGYW	NRKTDGNGWQ	ADPEHPLLGL	FAVSNVSMTL	AFVGICALVH
20		70	80	90	100	110	120
	orf18-1.pep	YCFSGTVQVF	VFAALLKLYA	LKPVYWFVLO	FVLMAYAYVH	RCGIDRQPPS	TFGGSQRLRG
	orf18ng	YCFSGTVQVF	VFAALLKLYA	LKPVYWFVLO	FVLMAYAYVH	RCGIDRQPPS	TFGGSQRLRG
25		130	140	150	160	170	180
	orf18-1.pep	GLTAALMQVS	VLVLLSEIGRX				
	orf18ng	VLAAMLMOVA	VTAMLLAEIGRX				
30		190	200				

Based on this analysis, including the presence of several putative transmembrane domains in the
35 gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and
their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 13

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 103>:

40 1 ATGAAAACCC CACTCCTCAA GCCTCTGCTN ATTACCTCGC TTCCCGTTTT
 51 CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA
 101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CTTGTGCGAT
 151 TTGGACAACC NCNTGACCGG ACGGCTNAAA AACATCATCA CCACCGTCGC
 201 CCTGTTCAAC CTCTCCTCGC TCACGGCACA AAGCACCCCTC GGCACAGGGC
 251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CTT.CG.CTT CACCATTTTA
 301 GGCGCGNCG ...

This corresponds to the amino acid sequence <SEQ ID 104; ORF19>:

1 MKTPLLKPLL ITS LPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
51 LDNXXTGRLK NIITVALEF LSSLTAQSTL GTGLPFILAM TLMTXFTIL
101 GAX...

50 Further work revealed the complete nucleotide sequence <SEQ ID 105>:

1 ATGAAAACCC CACTCCTCAA GCCTCTGCTC ATTACCTCGC TTCCCGTTTT
51 CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA
101 AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CTTGTGCGAT
151 TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCA CCACCGTCGC

201 CCTGTTCACC CTCTCCTCGC TCACGGCACA AAGCACCCCTC GGCACAGGGC
 251 TGCCCTTCAT CCTCGCCATG ACCCTGATGA CCTTCGGCTT CACCATTTTA
 301 GCGCGGTCG GGCTCAAATA CCGCACCTTC GCCTTCGGTG CACTCGCCGT
 351 CGCCACCTAC ACCACACTTA CCTACACCCC CGAAACCTAC TGGCTGACCA
 401 ACCCCTTCAT GATTTTATGC GGCACCGTAC TGTACAGCAC CGCCATCCTC
 451 CTGTTCCAAA TCGTCTCGCC CCACCGCCCC GTCCAAGAAA GCGTCGCCAA
 501 CGCCTACGAC GCACTCGGCG GCTACCTCGA AGCCAAAGCC GACTTCTTCG
 551 ACCCGATGA GGCAGCCTGG ATAGGCAACC GCCACATCGA CCTCGCCATG
 601 AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT
 651 TTACCGCCTT CGCGGCAAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC
 701 GTTACTACTT TGCCGCCCAA GACATACACG AACGCATCAG CTCGCCAC
 751 GTCGATTATC AGGAAATGTC CGAAAAATTC AAAAACACCG ACATCATCTT
 801 CGCATCCAC CGCCTGCTCG AAATGCAGGG ACAAGCCTGC CGCAACACCG
 851 CCCAAGCCCT GCGCGCAAGC AAAGACTACG TTTACAGCAA ACGCCTCGGC
 901 CGCGCCATCG AAGGCTGCGG CCAATCGCTG CGCCTCCTTT CAGACAGCAA
 951 CGACAGTCCC GACATCCGCC ACCTGCGCCG CTTTCTCGAC AACCTCGGCA
 1001 GCGTCGACCA GCAGTTCCGC CAACTCCAGC ACAACGGCCT GCAGGCAGAA
 1051 AACGACCGCA TGGGCGACAC CCGCATCGCC GCCCTCGAAA CCAGCAGCT
 1101 CAAAAACACC TGGCAGGCAA TCCGTCCGCA GCTAAACCTC GAATCAGGCG
 1151 TATTCGCCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC
 1201 ATCGTCGAAG CCCTCAACCT CAACCTCGGC TACTGGATAC TACTGACCGC
 1251 CTTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAGC CGCGTCCGCC
 1301 AGCGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
 1351 TACTTCACCC CGTCTGTGTA AACCAAACCTC TGGATTGTCA TCGCCAGTAC
 1401 CACCCTCTTT TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
 1451 TCATTACCAT TCAAGCCCTG ACCAGCCTCT CCCTCGCAGG TTTGGAGCTA
 1501 TAGCGCGCCA TGCCCGTACG CATCATCGAC ACCATTATCG GCGCATCCCT
 1551 TGCTGCGCG GCACTCAGCT ACCTGTGGCC AGACTGGAAG TACCTCACGC
 1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAACGGTGC CTATCTCGAA
 1651 AAAATCACCG AACGCCCTCA AAGCGGCGAA ACCGGCGACG ACGTCGAATA
 1701 CCGCGCCACC CGCCGCCCGC CCCACGAACA CACCGCCGCC CTCAGCAGCA
 1751 CCCTTTCGCA CATGAGCAGC GAACCGGCAA AATTCGCCGA CAGCCTGCAA
 1801 CCCGGCTTTA CCTGCTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC
 1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
 1901 TTACCGCACA GTTCCACCTC GCGCGCGAAC ACACCGCCCA CATCTTCCAA
 1951 CACCTGCCCG AAACCGAACC CGACGACTTT CAGACAGCAC TGGATACACT
 2001 CGCGGCGGAA CTCGACACCC TCCGCACCCA CAGCAGCGGA ACACAAAGCC
 2051 ACATCCTCCT CCAACAGCTC CAACTCATCG CCCGACAGCT CGAACCCCTAC
 2101 TACCGCGCCT ACCGCCAAAT TCCGCACAGG CAGCCCCAAA ATGCAGCCTG
 2151 A

This corresponds to the amino acid sequence <SEQ ID 106; ORF19-1>:

1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
 51 LDNRLTGRLE NIITVALEF LSSLTAQSTL GTGLFFILAM TLMTEGFTIL
 101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAIL
 151 LFQIVLPHRP VQESVANAYD ALGGYLEAKA DFFDPEAAW IGNRHIDLAM
 201 SNTGVITAFN QCRSALFYRL RGKRRHPRTA KMLRYFQAQ DIHERISSAH
 251 VDYQEMSEKF KNTDIIFRIH RLLEMGGQAC RNTAQALRAS KDYVYSKRLG
 301 RAIEGCRQSL RLLSDSNDSF DIRHLRRLD NLGSVDQQFR QLQHNGLQAE
 351 NDRMGDTRIA ALETSSLKNT WQAIRPOLNL ESGVFRHAVR LSLVVAACCT
 401 IVEALNINLG YWILLTALFV CQPNYATKS RVRQRIAGTV LGVIVGSLVP
 451 YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITIQAL TSLSLAGLDV
 501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLETAAL AVCSNGAYLE
 551 KITERLKSGE TGDDVEYRAT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ
 601 PGFTLLKTYG ALTGYSALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
 651 HLPETEPDDF QDALDTRGE LDTLRTHSSG TQSHILLOQL QLIARQLEPY
 701 YRAYRQIPHR QPQNAA*

Computer analysis of this amino acid sequence gave the following results:

Homology with predicted transmembrane protein YHFK of *H. influenzae* (accession number P44289)

ORF19 and YHFK proteins show 45% aa identity in 97 aa overlap:

60 orf19 6 LKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLKNIITT 65
 L +I+++PVF +V AA +W +MP +LGIIAGGLVDLDN TGRLKN+ T
 YHFK 5 LNAKVISTIPVFIADVNIAAVGIWFFDISSQSMPLILGIIAGGLVDLDNRLTGRLKNVFTT 64

orf19 66 VAFTLSSSLTAQSTLGTGLPFILAMTMTXXFTILGA 102
 + F++SS Q +G + +I+ MT++T FT++GA
 YHFK 65 LIAFSISSFIVQLHIGKPIQYIVLMTVLTFFIETMIGA 101

5 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF19 shows 92.2% identity over a 102aa overlap with an ORF (ORF19a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
10	orf19.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLPFVLGIIAGGLVDLDNXXTGRLK					
	orf19a	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLPFVLGIIAGGLVDLDNRLTGRLK					
		10	20	30	40	50	60
15	orf19.pep	NIIITVALFTLSSSLTAQSTLGTGLPFILAMTMTXXFTILGAX					
	orf19a	NIIATVALFTLSSSLVAQSTLGTGLPFILAMTMTFGFTIMGAVGLKYRTFAFGALAVATY					
		70	80	90	100	110	120
20	orf19a	TTLTYTPETYWLTNPFMILCGTVLYSTAILFQIILPHRPVQENVANAYEALGSYLEAKA					
		130	140	150	160	170	180

The complete length ORF19a nucleotide sequence <SEQ ID 107> is:

1	ATGAAAACCC	CACCCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
25	51	CGCCAGTGTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTG
	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCTGGCGG
	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG
	201	CCTGTTCAAC	CTCTCCTCAC	TTGTGCGGCA	AAGCACCTTC
	251	TGCCATTGAT	CCTCGCATAT	ACCCTGATGA	CTTTCGGCTT
	301	GGCGCGGTCG	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG
30	351	CGCCACCTAC	ACCACACTTA	CCTACACCCC	CGAAACCTAC
	401	ACCCCTTTAT	GATTCTGTGC	GGAACCGTAC	TGTACAGCAC
	451	CTGTTCCAAA	TCATCCTGCC	CCACCGCCCC	GTTCAAGAAA
	501	CGCCTACGAA	GCACTCGGCA	GCTACCTCGA	AGCCAAAGCC
	551	ATCCCGACGA	AGCCGAATGG	ATAGGCAACC	GCCACATCGA
35	601	AGCAACACCG	GCGTCATCAC	CGCCTTCAAC	CAATGCCGTT
	651	TTACCGCCTT	CGCGGCAAAC	ACCGCCACCC	GCGCACCGCC
	701	GCTACTACTT	CGCCGCCCAA	GACATACACG	AACGCATCAG
	751	GTCGACTACC	AAGAGATGTC	CGAAAAATTC	AAAAACACCG
	801	CCGCATCCAC	CGCCTGCTCG	AAATGCAGGG	ACAAGCCTGC
40	851	CCCAAGCCCT	GCGCGCAAGC	AAAGACTACG	TTTACAGCAA
	901	CGCGCCATCG	AAGGCTGCGG	CCAATCGCTG	CGCCTCCTTT
	951	CGACAATCCC	GACATCCGCC	ACCTGCGCGG	CCTTCTCGAC
	1001	GCGTCGACCA	GCAGTTCCGC	CAACTCCAGC	ACAACGGCCT
	1051	AACGACCGCA	TGGGCGACAC	CCGCATCGCC	GCCCTCGAAA
45	1101	CAAAAACACC	TGGCAGGCAA	TCCGTCCGCA	GCTAAACCTC
	1151	TATTCCGCCA	TGCCGTCCGC	CTGTCCCTTG	TCGTTGCCGC
	1201	ATCGTCGAAG	CCCTCAACCT	CAACCTCGGC	TACTGGATAC
	1251	CCTTTTCGTC	TGCCAACCCT	ACTACACCGC	CACCAAAAGC
	1301	AGCGCATCGC	CGGCACCGTA	CTCGGCGTAA	TCGTCGGCTC
50	1351	TACTTTACCC	CCTCCGTCTG	AACCAAACTC	TGGATCGTCA
	1401	CACCCTCTTT	TTTATGACCC	GCACCTACAA	ATACAGCTTC
	1451	TCATCACCAT	TCAAGCCCTG	ACCAGCCTCT	CCCTCGCAGG
	1501	TACGCGGCCA	TGCCCGTACG	CATCATCGAC	ACCATTATCG
	1551	TGCCTGGGCG	GCAGTCAGCT	ACCTGTGGCC	AGACTGGAAA
55	1601	TCGAAACGCAC	CGCCGCCCTT	GCCGTATGCA	GCAACGGCGC
	1651	AAAATCACCG	AACGCCTCAA	AAGCGCGGAA	ACCGGCGACG
	1701	CCGCGCCACC	CGCCGCCCGG	CCCACGAACA	CACCGCCGCC
	1751	CCCTTTCCGA	CATGAGCAGC	GAACCCGCAA	AATTCGCCGA
	1801	CCCGGCTTTA	CCCTGCTCAA	AACCGGCTAC	GCCCTGACCG
60	1851	CGCCCTCGGC	GCATACCGCA	GCGAAATGCA	CGAAGAATGC
	1901	TTACCGCACA	GTTCCACCTC	GCCGCCGAAC	ACACCGCCCA
	1951	CACCTGCCCC	AAACCGAACC	CGACGACTTT	CAGACAGCAC
	2001	GCGCGGCGAA	CTCGACACCC	TCCGCACCCA	CAGCAGCGGA
	2051	ACATCCTCCT	CCAACAGCTC	CAACTCATCG	CCCGGCAGCT
65	2101	TACCGCGCCT	ACCGACAAAT	TCCGCACAGG	CAGCCCCAAA
	2151	A			

This encodes a protein having amino acid sequence <SEQ ID 108>:

```

      1 MKTPPLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
    51 LDNRLTGRLL NIIATVALFT LSSLVAQSTL GTGLPFILAM TLMTFGFTIM
    101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAIL
    151 LFQIILPHRP VQENVANAYE ALGSYLEAKA DFFDPDEAEW IGNRHIDLAM
    201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYFFAAQ DIHERISSAH
    251 VDYQEMSEKF KNTDIIFRIH RLLEMQGOAC RNTAQALRAS KDYVYSKRLG
    301 RAIEGCRQSL RLLSDSNDNP DIRHLRRLD NLGSVDQOFR QLQHNGLQAE
    351 NDRMGDTRIA ALETGSLKNT WQAIRPOLNL ESGVFRHAVR LSLVVAAC
    401 IVEALNINLG YWILLTALFV CQPNYTATKS RVRQRIAGTV LGVIVGSLVP
    451 YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITIQUAL TSLSLAGLDV
    501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTERTAL AVCSNGAYLE
    551 KITERLKSGE TGDDVEYRAT RRAHEHTAA LSSTLSDMSS EPAKFADSLQ
    601 PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAETHAHIFQ
    651 HLPETEPDDF QTALDTLRGE LDTLRTHSSG TQSHILLQQL QLIARQLEPY
    701 YRAYRQIPHR QPQNAA*
  
```

ORF19a and ORF19-1 show 98.3% identity in 716 aa overlap:

		10	20	30	40	50	60
20	orf19a.pep	MKTPPLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIIAGGLVD	LDNRLTGRLL
	orf19-1	MKTPPLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIIAGGLVD	LDNRLTGRLL
		70	80	90	100	110	120
25	orf19a.pep	NIIATVALFT	LSSLVAQSTL	GTGLPFILAM	TLMTFGFTIM	GAVGLKYRTF	AFGALAVATY
	orf19-1	NIIATVALFT	LSSLVAQSTL	GTGLPFILAM	TLMTFGFTIM	GAVGLKYRTF	AFGALAVATY
		130	140	150	160	170	180
30	orf19a.pep	TTLTYTPETY	WLTNPFMILC	CGTVLYSTAIL	LFQIILPHRP	VQENVANAYE	ALGSYLEAKA
	orf19-1	TTLTYTPETY	WLTNPFMILC	CGTVLYSTAIL	LFQIILPHRP	VQENVANAYE	ALGSYLEAKA
		190	200	210	220	230	240
35	orf19a.pep	DFFDPDEAEW	IGNRHIDLAM	SNTGVITAFN	QCRSALFYRL	RGKHRHPRTA	KMLRYFFAAQ
	orf19-1	DFFDPDEAEW	IGNRHIDLAM	SNTGVITAFN	QCRSALFYRL	RGKHRHPRTA	KMLRYFFAAQ
		250	260	270	280	290	300
40	orf19a.pep	DIHERISSAH	VQEMSEKF	KNTDIIFRIH	RLLEMQGOAC	RNTAQALRAS	KDYVYSKRLG
	orf19-1	DIHERISSAH	VQEMSEKF	KNTDIIFRIH	RLLEMQGOAC	RNTAQALRAS	KDYVYSKRLG
		310	320	330	340	350	360
45	orf19a.pep	RAIEGCRQSL	RLLSDSNDNP	DIRHLRRLD	NLGSVDQOFR	QLQHNGLQAE	ENDRMGDTRIA
	orf19-1	RAIEGCRQSL	RLLSDSNDNP	DIRHLRRLD	NLGSVDQOFR	QLQHNGLQAE	ENDRMGDTRIA
		370	380	390	400	410	420
50	orf19a.pep	ALETGSLKNT	WQAIRPOLNL	ESGVFRHAVR	LSLVVAAC	IVEALNINLG	YWILLTALFV
	orf19-1	ALETGSLKNT	WQAIRPOLNL	ESGVFRHAVR	LSLVVAAC	IVEALNINLG	YWILLTALFV
		430	440	450	460	470	480
55	orf19a.pep	CQPNYTATKS	RVQRRIAGTV	LVGIVGSLVP	YFTPSVETKL	WIVIASTTLF	FMTRTYKYSF
	orf19-1	CQPNYTATKS	RVQRRIAGTV	LVGIVGSLVP	YFTPSVETKL	WIVIASTTLF	FMTRTYKYSF
		490	500	510	520	530	540
60	orf19a.pep	STFFITIQUAL	TSLSLAGLDV	YAAMPVRIID	TIIGASLAWA	AVSYLWPDWK	YLTERTAL
	orf19-1	STFFITIQUAL	TSLSLAGLDV	YAAMPVRIID	TIIGASLAWA	AVSYLWPDWK	YLTERTAL

-112-

5	orf19-1	STFFITIQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAAL	490	500	510	520	530	540
	orf19a.pep	AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSSEPAKFADSLQ	550	560	570	580	590	600
10	orf19-1	PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPETEPDDF	610	620	630	640	650	660
	orf19a.pep	PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPETEPDDF	610	620	630	640	650	660
15	orf19-1	QTALDTLRGELDTLRTHSSGTQSHILLQQLQLIARQLEPYRAYRQIPHRQPQNAAX	670	680	690	700	710	
	orf19a.pep	QTALDTLRGELDTLRTHSSGTQSHILLQQLQLIARQLEPYRAYRQIPHRQPQNAAX	670	680	690	700	710	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF19 shows 95.1% identity over a 102aa overlap with a predicted ORF (ORF19.ng) from *N. gonorrhoeae*:

25	orf19.pep	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLK	60
	orf19ng	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK	60
30	orf19.pep	NIIITVALEFTLSSLTQSTLGTGLPFILAMTLMTXXFTILGAX	103
	orf19ng	NIIATVALEFTLSSLTQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTAFAGALAVATY	120

An ORF19ng nucleotide sequence <SEQ ID 109> is predicted to encode a protein having amino acid sequence <SEQ ID 110>:

35	1	MKTPLLKPLL	ITSLPVFASV	FTAASIVWQL	GEPKLAMPFV	LGIIAGGLVD
	51	LDNRLTGRLK	NIIATVALEFT	LSSLTQSTL	GTGLPFILAM	TLMTXXFTIL
40	101	GAVGLKYRTF	AFGALAVATY	TTLTYTPETY	WLTNPFMILC	GTVLYSTAII
	151	LFQIILPHRP	VOESVANAYE	ALGGYLEAKA	DFDPPDEAAW	IGNRHIDLAM
45	201	NTSGVITAFN	QCRSALFYRL	RGKRRHPRTA	KMLRYFFAAQ	DIHERISSAH
	251	VDYQEMSEKF	KNTDIIIFRI	RLEMOGQAC	RNTAQAIRSG	KDYVYSKRLG
50	301	RAIEGCRQSL	RLSDGNDSP	DIRHLSRLLD	NLGSVDQQFR	QLRHSDSPA
	351	NDRMGDTRIA	ALETGSFKNT	*		

Further work revealed the complete nucleotide sequence <SEQ ID 111>:

45	1	ATGAAAACCC	CACTCCTCAA	GCCTCTGCTC	ATTACCTCGC	TTCCCGTTTT
	51	CGCCAGTGTC	TTTACCGCCG	CCTCCATCGT	CTGGCAGCTA	GGCGAACCCA
50	101	AGCTCGCCAT	GCCCTTCGTA	CTCGGCATCA	TCGCCGGCGG	CCTGGTCGAT
	151	TTGGACAACC	GCCTGACCGG	ACGGCTGAAA	AACATCATCG	CCACCGTCGC
55	201	CCTGTTTACC	CTCTCCTCGC	TCACGGCGCA	AAGCACCCTC	GGCACAGGGC
	251	TGCCCTTCAT	CCTCGCCATG	ACCCTGATGA	CCTTCGGCTT	TACCATTTTA
60	301	GGCGCGGTCG	GGCTGAAATA	CCGCACCTTC	GCCTTCGGCG	CACTCGCCGT
	351	CGCCACCTAC	ACCACGCTTA	CCTACACCCC	CGAAACCTAC	TGGCTGACCA
65	401	ACCCCTTCAT	GATTTTATGC	GGCACCGTAC	TGTACAGCAC	CGCCATCATC
	451	CTGTTCCAAA	TCATCTGCC	CCACCGCCCC	GTCCAAGAAA	GCGTCGCCAA
70	501	TGCCTACGAA	GCACTCGGCG	GCTACCTCGA	AGCCAAAGCC	GACTTCTTCG
	551	ACCCCGATGA	GGCAGCTGG	ATAGCAACC	GCCACATCGA	CCTCGCCATG
75	601	AGCAACACCG	CGCTCATCAC	CGCCTTCAAC	CAATGCCGTT	CCGCCCTGTT
	651	TTACCGTTTG	CGCGGCAAAC	ACCGCCACCC	GCGCACCGCC	AAAAATGCTG
80	701	GCTACTACTT	CGCCGCCCAA	GACATCCACG	AACGCATCAG	CTCCGCCAC
	751	GTCGACTACC	AAGAGATGTC	CGAAAATTTC	AAAAACACCG	ACATCATCTT
85	801	CCGCATCCGC	CGCCTGCTCG	AAATGCAGGG	GCAGGCGTGC	CGCAACACCG
	851	CCCAAGCCAT	CCGGTCGGGC	AAAGACTAcg	tTTACAGCAA	ACGCCTCGGA
90	901	CGCGCCATcg	aaggctgCCG	CCAGTCGCTg	cgcctCCTTt	cagacggca
	951	CGACAGTCCC	GACATCCGCC	ACCTGAGccg	CCTTCTCGAC	AACCTCGgca

-113-

1001 GCGTcgacca gcagtTCcgc caactCCGAC ACAGcgactC CCCCgcgaa
 1051 Aacgaccgca tggcgacac CCGCATCGCC GCCctcgaaa ccggcagctT
 1101 caaaaaCacc tggcaggCAA TCCGTCCGCA gctgaaCCTC GAATCatgCG
 1151 TATTCCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC
 5 1201 ATCGTCgaag cCCTAACCT CAACCTCGGC TACTGGATAC TGCTGACCGC
 1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTGTACC
 1301 AACGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
 1351 TACTTCACCC CCTCCGTGCG AACCAAATC TGGATTGTCA TCGCCGGTAC
 1401 CACCCTGTTC TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
 10 1451 TCATCACCAT TCAGGCACTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA
 1501 TACGCCGCCA TGCCGTGCG CATCategaC ACCATTATCG GCGCATCCCT
 1551 TGCCTGGGCG GCGGTGAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC
 1601 TCGAACGCAC CGCGCCCTT GCCGTATGCA GCAGCGGCAC ATACCTCCAA
 1651 AAAATTGCCG AACGCCTCAA AACCGCGGAA ACCGGCGAGC ACATAGAATA
 15 1701 CCGCATCACC CGCCGCCGCG CCCACGAACA CACCGCGGCC CTCAGCAGCA
 1751 CCCTTTCCGA CATGAGCAGC GAACCGCAA AATTCGCCA CAGCCTGCAA
 1801 CCGCGCTTTA CCCTGCTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC
 1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCGGACT
 1901 TTACCGCACA GTTCCACCTT GCCGCCGAAC ACACCGCCCA CATCTTCCAA
 20 1951 CACCTGCCCG ACATGGGACC CGACGACTTT CAGACGGCAT TGGATACACT
 2001 GCGCGCGCAA CTCGGCACCC TCCGCACCCG CAGCAGCGGA ACACAAAGCC
 2051 ACATCCTCCT CCAACAGCTC CAACTCATCG CcgcGCAACT CGAACCCCTAC
 2101 TACCGCGCCT ACCGACAAAT TCCGCACAGG CAGCCCCAAA ACGCAGCCTG
 2151 A

25 This corresponds to the amino acid sequence <SEQ ID 112; ORF19ng-1>:

1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPLKAMPFV LGIIAGGLVD
 51 LDNRLTGRK NIIATVALFT LSSLTAQSTL GTGLPFILAM TLMTEGFTIL
 101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAI
 151 LFQIILPHRP VQESVANAYE ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM
 201 SNTGVITAFN QCRSALFYRL RGKXHRHPTA KMLRYFAAQ DIHERISSAH
 251 VDYQEMSEKF KNTDIIIFRIR RLEMQGQAC RNTAQAIRSG KDYVYSKRLG
 301 RAIEGCRQSL RLLSDGNDSP DIRHLSRLLD NLGSVDQQFR QLRHSDSPA
 351 NDRMGDTRIA ALETGSFKNT WQAIRPQLNL ESCVFRHAVR LSLVVAAGT
 401 IVEALNLNLG YWILLTALFV QPNYTATKS RVYQRIAGTV LGVIVGSLVP
 35 451 YFTPSVETKL WIVIAGTTLF FMTRTYKYSF STFFITIQAL TSLSLAGLDV
 501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLETAAL AVCSSSTYLO
 551 KIAERLKTGE TGDDIEYRIT RRAHEHTAA LSSTLSMSS EPAKFADSLQ
 601 PGFTLLKTGY ALTGYSALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
 651 HLPDMGPDDF QTALDILRGE LGTLRTRSSG TQSHILLQOL QLIARQLEPY
 40 701 YRAYRQIPHR QPQNAA*

ORF19ng-1 and ORF19-1 show 95.5% identity in 716 aa overlap:

		10	20	30	40	50	60
orf19-1.pep		MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPLKAMPFVLGIIAGGLVDLDNRLTGRK					
45	orf19ng-1	MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPLKAMPFVLGIIAGGLVDLDNRLTGRK					
		10	20	30	40	50	60
	orf19-1.pep	NIIITVALEFTLSSLTAQSTLGTGLPFILAMTLMTEGFTILGAVGLKYRTFAFGALAVATY					
50	orf19ng-1	NIIATVALFTLSSLTAQSTLGTGLPFILAMTLMTEGFTILGAVGLKYRTFAFGALAVATY					
		70	80	90	100	110	120
	orf19-1.pep	TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIIVLPHRPVQESVANAYDALGGYLEAKA					
	orf19ng-1	TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIILPHRPVQESVANAYEALGGYLEAKA					
		130	140	150	160	170	180
55	orf19-1.pep	DFFDPDEAAWIGNRHIDLAMSNITGVITAFNQCRSALFYRLRGKXHRHPTAKMLRYFAAQ					
	orf19ng-1	DFFDPDEAAWIGNRHIDLAMSNITGVITAFNQCRSALFYRLRGKXHRHPTAKMLRYFAAQ					
		190	200	210	220	230	240
60	orf19-1.pep	DIHERISSAHVDYQEMSEKFNTDIIIFRIHRLLEMQGQACRNTAQALRASKDYVYSKRLG					
	orf19ng-1	DIHERISSAHVDYQEMSEKFNTDIIIFRIHRLLEMQGQACRNTAQALRASKDYVYSKRLG					
		250	260	270	280	290	300
65	orf19-1.pep						

-114-

5	orf19ng-1	 DIHERISSAHVDYQEMSEKFKNTDIIIFRIRRLLEMGGQACRNTAQAIRSGKDYVYSKRLG	250	260	270	280	290	300
	orf19-1.pep	310 320 330 340 350 360 RAIEGCRQSLRLLSDSNDSPDIRHLRRLLDNLGSDVQQFRQLQHNGLQAENDRMGDTRIA						
10	orf19ng-1	 RAIEGCRQSLRLLSDGNDSPDIRHLSRLLLDNLGSDVQQFRQLRHSDSPAENDRMGDTRIA	310	320	330	340	350	360
	orf19-1.pep	370 380 390 400 410 420 ALETSSLKNTWQAIRPQLNLESGVFRHAVRLSLVVAACTIVEALNINLGYWILLTALFV						
15	orf19ng-1	 ALETGSFKNTWQAIRPQLNLESCVFRHAVRLSLVVAACTIVEALNINLGYWILLTALFV	370	380	390	400	410	420
	orf19-1.pep	430 440 450 460 470 480 CQPNYTATKSRVQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF						
20	orf19ng-1	 CQPNYTATKSRVQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF	430	440	450	460	470	480
	orf19-1.pep	490 500 510 520 530 540 STFFITIQUALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLETAAL						
25	orf19ng-1	 STFFITIQUALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLETAAL	490	500	510	520	530	540
	orf19-1.pep	550 560 570 580 590 600 AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSEPAKFADSLQ						
30	orf19ng-1	 AVCSSGYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADSLQ	550	560	570	580	590	600
	orf19-1.pep	610 620 630 640 650 660 PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPETEPDDF						
35	orf19ng-1	 PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPDMGPDDF	610	620	630	640	650	660
	orf19-1.pep	670 680 690 700 710 QTALDTLRGELDTLRTHSSGTQSHILLQQLLIARQLEPYRAYRQIPHROPQNAAX						
40	orf19ng-1	 QTALDTLRGELGTLRTRSSGTQSHILLQQLLIARQLEPYRAYRQIPHROPQNAAX	670	680	690	700	710	
	orf19-1.pep							

In addition, ORF19ng-1 shows significant homology to a hypothetical gonococcal protein previously entered in the databases:

50	sp O33369 YOR2_NEIGO HYPOTHETICAL 45.5 KD PROTEIN (ORF2) gnl PID e1154438 (AJ002423) hypothetical protein [Neisseria gonorrh] Length = 417	
	Score = 1512 (705.6 bits), Expect = 5.3e-203, P = 5.3e-203 Identities = 301/326 (92%), Positives = 306/326 (93%)	
55	Query:	307 RQSLRLLSDGNDSPDIRHLSRLLLDNLGSDVQQFRQLRHSDSPAENDRMGDTRIAALETGS 366
	Sbjct:	1 RQSLRLLSDGNDSPDIRHLSRLLLDNLGSDVQQFRQLRHSDSPAENDRMGDTRIAALETGS 60
60	Query:	367 FKNTWQAIRPQLNLESCVFRHAVRLSLVVAACTIVEALNINLGYWILLTALFVCQPNYT 426
	Sbjct:	61 FKNTWQAIRPQLNLESGVFRHAVRLSLVVAACTIVEALNINLGYWILLTRLFVCQPNYT 120
65	Query:	427 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSFSTFFIT 486
	Sbjct:	121 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSFSTFFIT 180
	Query:	487 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLETAALAVCSSG 546
	Sbjct:	181 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLETAALAVCSSG 240

Query: 547 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSMSSEPAKFADSLOPGFTLL 606
 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSMSSEPAKFAD+ P
 Sbjct: 241 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSMSSEPAKFADTCNPALPCS 300

5 Query: 607 KTGyALTGYISALGAYRSEMHEECSP 632
 K ALTGYISALG ++ + +P
 Sbjct: 301 KPATALTGYISALGHTAAKCTKNAAP 326

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein (the first of which is also seen in the meningococcal protein), and on homology
 10 with the YHFK protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 14

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 113>:

```

15      1 ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTG
      51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTGCGCG
     101 CGGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAAC T GCCCAACCTG
     151 CTTCGCCGCG TGTTCGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
     201 TTTGGCGGAA TACAAGGAAA CGCGTTCAA AGAGGCGG.C GAAGCCTTTA
     251 TCCGCCATGT GCGGGGATG CTGTCGTTG TACTGGTTAT CGTTACCGCG
     301 CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTCCG CACCCGAGTT
     351 TTGCCCAAGA TGCCGACAAA TTTCAGTCT CCATCGATT GCTGCGGATT
     401 ACGTTTCCTT ATATATTATT GATTTCCTG TCTTCATTG TCGGCTCGGT
     451 ACTCAATTCT TATCATAAGT TCGGCATTCC GCGTTCACG CCAC.GTTTC
     501 TGAACGTGTC GTTTATCGTA TTCGCGCTGT TTTTCGTGCC GTATTTTCAT
     551 CCGCCCGTTA CCGCGCGyGGC GTGGGCGGTC TTTGTGCGCG GCATTTTGCA
     601 ACTCGmTTC CAACTGCCCT GGCTGGCGAA ACTGGGCTTT TTGAAACTGC
     651 CCAAACTGAG TTTCAAAGAT GCGGCGGTCA ACCGCGTGAT GAAACAGATG
     701 GCGCCTGCga TTTTgGGCGT GAgCGTGGCG CAGGTTTCTT TGGTGATCAA
     751 CACGATTTt GCGTCTTATC TGCAATCGGG CAGCGTTTCA TGGATGTATT
     801 ACGCCGACCG CATGATGGAG CTGCCAGCG GCGTGTGGG GCGGCGACTC
     851 GGTACGATTT TGCTGCCGAC TTTGTCAAAA CACTCGCAA ACCaAGATAC
     901 GGaACAGTTT TCCGCCCTGC TCGACTGGG TTTGCGCTG TGCATGCTgc
     951 TGACGCTGCC GCGGgcGGTC GGACTGGCGG TGTTGTCGTT cCCgCtGGTG
    1001 GCGACGCTGT TTATGTACCG CGwATTACG CTGTTTGACG CGCAGATGAC
    1051 GCAACACGCG CTGATTGCCT ATTCTTTCGG TTTAATCGGC TTAATCATGA
    1101 TTAAAGTGTT GGCACCCGGC TTCTATGCGC GGCAAAACAT CAAwAmGCC
    1151 GTCAAAATCG CCATCTTCAC GCTCATCTGC mCGCAGTTGA TGAACCTTGs
    1201 CTTTayCGGC CCACTrAAC rCagTCGGAC TTTGCTTGC CATCGGTCTG
    1251 GGCGCGTGTA TCAATGCCCG ATTGTTGTTT TACCTGTTGC GCAGACACGG
    1301 TATTTACCAA CCTGG.CAAG GGTGGGCAG CGTTCTT.AG CAAAAATGCT
    1351 GcTCTCGCTC GCCGTGA
  
```

This corresponds to the amino acid sequence <SEQ ID 114; ORF20>:

```

45      1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL
      51 LRRVFAEGAF AQAFVPILAE YKETRSKEAX EAFIRHVAGM LSEVLVIVTA
     101 LGILAAPWVI YVSAPSFAQD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
     151 LNSYHKFGIP AFTPXFLNVS FIVFALFFVP YFDPVPTAXA WAVFVGGILQ
     201 LXFLPLWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN
     251 TIFASYLQSG SVSWMYADR MMELPSGLV AALGTILLPT LSKHSANQDT
     301 EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR XFTLFDAQMT
     351 QHALIAYSFG LIGLIMIKVL APGFYARQNI XXPVKIAIFT LICXQLMNLX
     401 FXGPLXXIGL SLAIGLGACI NAGLLFYLLR RHGIYQPXQG LGSVLXQKCC
     451 SRSP*
  
```

These sequences were elaborated, and the complete DNA sequence <SEQ ID 115> is:

```

55      1 ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTG
      51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTGCGCG
  
```

101 CGGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAACCT GCCCAACCTG
 151 CTTCGCCGCG TGTTCGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
 201 TTTGGCGGAA TACAAGGAAA CGCGTTCAAA AGAGGCGGCG GAGGCTTTTA
 251 TCCGCCATGT GCGGGGATG CTGTCGTTTG TACTGGTTAT CGTTACCGCG
 301 CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGGTTT
 351 TGCCCAAGAT GCCGACAAAT TTCAGCTCTC CATCGATTG CTGCGGATTA
 401 CGTTTCCTTA TATATTATTG ATTTCCCTGT CTTCAATTTGT CGGCTCGGTA
 451 CTCGAATCTT ATCATAAGTT CGGCATTCCG GCGTTTACGC CCACGTTTCT
 501 GAACGTGTCG TTTATCGTAT TCGCGCTGTT TTTCTGCGC TATTTGCGATC
 551 CGCCCGTTAC CGCGCTGGCG TGGGCGGTCT TTGTGCGCG CATTTTGCAA
 601 CTCGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC
 651 CAAACTGAGT TTCAAAGATG CGGCGGTCAA CCGCGTGATG AAACAGATGG
 701 CGCTGCGAT TTTGGGCGTG AGCGTGGCGC AGGTTTCTTT GGTGATCAAC
 751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTATTA
 801 CGCCGACCGC ATGATGGAGC TGCCGAGCGG CGTGTGGGG GCGGCACTCG
 851 GTACGATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
 901 GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCTGT GCATGCTGT
 951 GACGCTGCCG GCGGCGGTCT GACTGGCGGT GTTGTGTTT CCGTGTGGT
 1001 CGACGCTGTT TATGTACCGC GAATTTACGC TGTGACGC GCAGATGACG
 1051 CAACACGCGC TGATTGCTTA TTCTTCGGT TTAATCGGT TAATCATGAT
 1101 TAAAGTGTG GCACCGGCT TCTATGCGC GCAAACATC AAAACGCCCG
 1151 TCAAATCGC CATCTTACG CTCATCTGCA CGCAGTTGAT GAACCTTGCC
 1201 TTTATCGGCC CACTGAAACA CGTCGGACTT TCGCTTGCCA TCGGTCTGGG
 1251 CGCGTGATC AATGCCGGAT TGTGTTTGA CCGTGTGCG AGACACGGTA
 1301 TTTACCAACC TGGCAAGGGT TGGGACGCGT TCTTAGCAA AATGCTGCTC
 1351 TCGCTCGCCG TGATGTGCGG CGGACTGTGG GCAGCGCAGG CTTACCTGCC
 1401 GTTTGAATGG GCGCACGCCG GCGGAATCG GAAAGCGGG CAGCTCTGCA
 1451 TCCTGATTGC CGTCGGCGGC GGAATGTATT TCGCATCACT GCGGCGTTTG
 1501 GGCTTCCGTC CGCGCCATT CAAACGCGTG GAAACTGA

30 This corresponds to the amino acid sequence <SEQ ID 116; ORF20-1>:

1 MNMLGALAKV GSITMVSRLV GFVRDVIAR AFGAGMATDA FFVAFKLPNL
 51 LRVFAEGAF AQAFVPILAE YKETRSKEAA EAFIRHVAGM LSFVLVIVTA
 101 LGILAAPWVI YVSAPGFAQD ADKFQLSIDL LRITFPYILL ISLSFVGSV
 151 LNSYHKFGIP AFTPTFLNVS FIVFALFFVP YFDPVVTALA WAVEVGGILO
 201 LGFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN
 251 TIFASYLQSG SVSWMYADR MMELPSGVLG AALGTILLPT LSKHSANQDT
 301 EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLEMYR EFTLFDAQMT
 351 QHALIAYSFG LIGLIMIKVL APGFYARONI KTPVKIAIFT LICTQLMNL
 401 FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL
 451 SLAVMCGGLW AAQAYLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL
 501 GFRPRHFKR V EN*

Computer analysis of this amino acid sequence gave the following results:

Homology with the MviN virulence factor of *S. typhimurium* (accession number P37169)

ORF20 and MviN proteins show 63% aa identity in 440aa overlap:

45 Orf20 1 MNMLGALAKVGSITMVSRLVGFVRDVIARAFAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 60
 MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDAFFVAFKLPNLLRR+FAEGAF
 MviN 14 MNLLKSLAAVSSMTMFSRVLGFARDAIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF 73
 50 Orf20 61 AQAFVPILAEYKETRSKEAXEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPGFAQD 120
 +AQAFVPILAEYK + +EA F+ +V+G+L+ L +VT G+LAAPWVI V+AP FA
 MviN 74 SQAFVPILAEYKSKQGEETRIFVAYVSGLLTLALAVTVAGMLAAPWVIMVTAPGFADT 133
 55 Orf20 121 ADKFQLSIDLLRITFPYILLISLSFVGSVLNLSYHKFGIPFTFPFLNVSFIVFALFFVP 180
 ADKF L+ LLRITFPYILLISL+S VG++LN+++F IPAF P FLN+S I FALF P
 MviN 134 ADKFALTQLLRITFPYILLISLASLVGAILNTWNRFIPAFAPTFLNISMIGFALFAAP 193
 60 Orf20 181 YFDPVVTAXAWAVFVGGILQLXFLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV 240
 YF+PPV A AWAV VGG+LQL +QLP+L K+G L LP+++F+D RV+KQM PAILGV
 MviN 194 YFNPPVLALAWAVTVGGVLQVYQLPYLKKIGMLVLPRIINFRDTGAMRVVKQMGPAILGV 253
 Orf20 241 SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT 300
 SV+Q+SL+INTIFAS+L SGSVSWMYADR+ME PSGVLG ALGTILLP+LSK A+ +
 MviN 254 SVSQISLIINTIFASFLASGVSWSMYADRLMEFPGSVLGAALGTILLPSLSKSFASGNH 313

Orf20 301 EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG 360
 +++ L+DWGLRLC LL LP+AV L +L+ PL +LF Y FT FDA MTQ ALIAYS G
 MviN 314 DEYCRIMDWGLRLCFLALPSAVALGILAKPLTVSLFQYGKFTAFDAAMTQRALIAYSFG 373

5 Orf20 361 LIGLIMIKVLAPGFYARONIXXPVKIAIFTLICXQLMNLXFXXXXXXXXXXXXXXXXXXCI 420
 LIGLI++KVLAPGFY+RQ+I PVKIAI TLI QLMNL F C+
 MviN 374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

10 Orf20 421 NAGLLFYLLRRHGIYQXPQG 440
 NA LL++ LR+ I+ P G
 MviN 434 NASLLYWQLRKQNIPTQPG 453

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF20 shows 93.5% identity over a 447aa overlap with an ORF (ORF20a) from strain A of *N.*

15 *meningitidis*:

		10	20	30	40	50	60
orf20.pep		MNMLGALAKVGS	SLTMVSRVLGF	VRDVIARAFG	GAGMATDAFF	VAFKLPNLLRR	VFAEGAF
orf20a		MNMLGALVKVGS	SLTMVSRVLGF	VRDVIARAFG	GAGMATDAFF	VAFKLPNLLRR	VFAEGAF
20		10	20	30	40	50	60
		70	80	90	100	110	120
orf20.pep		AQAFVPILA	EYKETSKEA	EAFIRHVAG	MLSFVLVIV	TALGILAAP	VWVIYVSAP
orf20a		AQAFVPILA	EYKETSKEA	EAFIRHVAG	MLSFVLVIV	TALGILAAP	VWVIYVSAP
25		70	80	90	100	110	120
		130	140	150	160	170	180
orf20.pep		ADKFQLSID	LLRITFPYI	LLISLSSFV	GSVLNSYHK	FGIPAFTPX	FLNVSFIVF
orf20a		ADKFQLSID	LLRITFPYI	LLISLSSFV	GSVLNSYHK	FGIPAFTPX	FLNVSFIVF
30		130	140	150	160	170	180
		190	200	210	220	230	240
orf20.pep		YFDPPTAXA	WAVFVGGI	LQLXFLPWL	AKLGLKLP	KLSPKLSFK	DAAVNRVM
orf20a		YFDPPTAXA	WAVFVGGI	LQLXFLPWL	AKLGLKLP	KLSPKLSFK	DAAVNRVM
35		190	200	210	220	230	240
		250	260	270	280	290	300
orf20.pep		SVAQVSLV	INTIFASYL	QSGSVSWM	YYADRMME	LPSGVLGA	ALGTILLPT
orf20a		SVAQVSLV	INTIFASYL	QSGSVSWM	YYADRMME	LPSGVLGA	ALGTILLPT
40		250	260	270	280	290	300
		310	320	330	340	350	360
orf20.pep		EQFSALLDW	GLRLCMLLT	LPAAVGLAV	LSFPLVATL	FMYRXFTLF	DAQMTQHAI
orf20a		EQFSALLDW	GLRLCMLLT	LPAAVGLAV	LSFPLVATL	FMYRXFTLF	DAQMTQHAI
50		310	320	330	340	350	360
		370	380	390	400	410	420
orf20.pep		LIGLIMIKV	LAPGFYAR	ONIXXPVKI	AIFTLICXQ	LMNLXFXG	PLXXIGLSL
orf20a		LIGLIMIKV	LAPGFYAR	ONIXXPVKI	AIFTLICXQ	LMNLXFXG	PLXXIGLSL
55		370	380	390	400	410	420
		430	440	450			
orf20.pep		NAGLLFYLL	RRHGIYQP	XQGLSVLX	QKCCSRSP		
orf20a		NAGLLFYLL	RRHGIYQP	XQGLSVLX	QKCCSRSP		
60		430	440	450	460	470	480

The complete length ORF20a nucleotide sequence <SEQ ID 117> is:

65 1 ATGAATATGC TGGGAGCTTT GGTAAGATC GGCAGCCTGA CGATGGTGTG
 51 GCGCGTTTGG GGATTTGTGC GCGATACGGT CATTGCGCGC GCATTGCGCG
 101 CAGGCATGGC GACGGATGCG TTCTTTGTGC CGTTCAAACCT GCCCAACCTG

151 CTTCCGCCGCG TGTTCGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
 201 TTTGGCGGAA TATAAGGAAA CGCGTTCTAA AGAGGCGACG GAGGCTTTTA
 251 TCCGCCATGT GCGCGGGATG CTGTCGTTTG TACTGGTCAT CGTTACCGCG
 301 CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTCCG CACCCGGTTT
 351 TGCCAAAGAT GCCGACAAAT TTCAGCTCTC TATCGATTG CTGCGGATTA
 401 CGTTTCCTTA TATCTTATTG ATTTCACTTT CCTCTTTGT CGGCTCGGTA
 451 CTCAATTCCT ATCATAAATT CAGCATTCCT GCGTTTACGC CCACGTTCTT
 501 GAACGTGTCG TTTATCGTAT TCGCGCTGTT TTTCGTGCGG TATTTCGATC
 551 CTCCCGTTAC CGCGCTGGCT TGGGCGGTTT TTGTCGCGCG CATTTTGCAA
 601 CTCGGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGTTTTT TGAACTGCC
 651 CAAACTGAGT TTCAAAGATG CGGCGGTCAA CCGCGTGATG AAACAGATGG
 701 CGCCTGCGAT TTGGGCGGTG AGCGTGGCGC AGATTCTTTT GGTGATCAAC
 751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTATTA
 801 CGCCGACCGC ATGATGGAAC TGCCCGGCGG CGTGCTGGGG GCGGCATCG
 851 GTACGATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
 901 GAACAGTTT CCGCCCTGCT CGACTGGGGT TTGCGCNTGT GCATGCTGCT
 951 GACGCTGCCG GCGGCGGTG GAATGGCGGT GTTGTGCTTC CCGCTGGTGG
 1001 CAACCTTGTT TATGTACCGA GAATTCACGC TGTTCGACG GCAGATGACG
 1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGTT TAATCATGAT
 1101 TAAAGTGTG GCGCCCGGCT TTTATGCGCG GCAAAACATC AAAACGCCCG
 1151 TCAAAATCGC CATCTTCACG CTCATTGCA CGCAGTTGAT GAACCTTGCC
 1201 TTTATCGGCC CACTGAAACA CGTCGGACTT TCGCTTGCCA TCGGTCTGGG
 1251 CGCGTGTATC AATGCCGGAT TGTGTTTGA CCTGTTGCGC AGACACGTA
 1301 TTTACCAACC TGGCAAGGGT TGGCAGCGT TCTTGGCAA AATGCTGCTC
 1351 TCGCTCGCCG TGATGGGAGG CGGCTGTAT GCCGCCCAA TCTGGCTGCC
 1401 GTTCGACTGG GCACACGCCG GCGGAATGCA AAAGGCCGCC CGGCTCTTCA
 1451 TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCACT GCGGCTTTG
 1501 GGCTCCGTC CGGCCATT TCAAACGCGT GAAAGCTGA

This encodes a protein having amino acid sequence <SEQ ID 118>:

30 1 MNMLGALVKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL
 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLVIVTA
 101 LGILAAPWVI YVSAPGFAKD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
 151 LNSYHKFSIP AFTPTFLNVS FIVFALFFVP YFDPVPTALA WAVFVGGLQ
 201 LGFLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQISLVIN
 251 TIFASYLQSG SVSWMYADR MMELPGGVLG AALGTILLPT LSKHSANQDT
 301 EQFSALLDWG LRXCMLLTLP AAVGMAVLSF PLVATLFMYR EFTLFDAQMT
 351 QHALIAYSFG LIGLIMIKVL APGFYARONI KTPVKIAIFT LICTOLMNL
 401 FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL
 451 SLAVMGGGLY AAQIWLFPDW AHAGGMQKAA RLFILIAVGG GLYFASLAAL
 501 GFRPRHFKRV ES*

ORF20a and ORF20-1 show 96.5% identity in 512 aa overlap:

		10	20	30	40	50	60
45	orf20a.pep	MNMLGALVKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL LRRVFAEGAF					
	orf20-1	MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL LRRVFAEGAF					
		10	20	30	40	50	60
50	orf20a.pep	AQAFVPILAE YKETRSKEATEAFIRHVAGM LSFVLVIVTALGILAAPWVI YVSAPGFAKD					
	orf20-1	AQAFVPILAE YKETRSKEAAEAFIRHVAGM LSFVLVIVTALGILAAPWVI YVSAPGFAQD					
		70	80	90	100	110	120
55	orf20a.pep	ADKFQLSIDL LRITFPYILL ISLSSFVGSV LNSYHKFSIP AFTPTFLNVS FIVFALFFVP					
	orf20-1	ADKFQLSIDL LRITFPYILL ISLSSFVGSV LNSYHKFGIP AFTPTFLNVS FIVFALFFVP					
		130	140	150	160	170	180
60	orf20a.pep	YFDPVPTALAWAVFVGGLQLGFLPWLAKLGLKLPKLSFKDAAVNRVMKQMAPAILGV					
	orf20-1	YFDPVPTALAWAVFVGGLQLGFLPWLAKLGLKLPKLSFKDAAVNRVMKQMAPAILGV					
		190	200	210	220	230	240
65	orf20a.pep	SVAQISLVINTIFASYLQSGSVSWMYADRMELPGGVLG AALGTILLPTLSKHSANQDT					

30 ORF20 shows 92.1% identity over a 454aa overlap with a predicted ORF (ORF20ng) from *N. gonorrhoeae*:

Accession	Gene	Protein	Length
35	orf20.pep	MNMLGALAKVGSGLTMVSRVLGFVRDVTIARAFAAGMATDAFFVAFKLPNLLRRVFAEGAF	60
	orf20ng	MNMLGALAKVGSGLTMVSRVLGFVRDVTIARAFAAGMATDAFFVAFKLPNLLRRVFAEGAF	60
40	orf20.pep	AQAFVPILAELYKETRSKEAEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPSFAQD	120
	orf20ng	AQAFVPILAELYKETRSKEATEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPGFTKD	120
45	orf20.pep	ADKFQLSIDLRLITFPYILLISLSSFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP	180
	orf20ng	ADKFQLSISLLRLITFPYILLISLSSFVGSILNSYHKFGIPAFTPFTFLNISFIVFALFFVP	180
50	orf20.pep	YFDPVPTAXAWAVFVGILQLXFLPWLAKLGLKLPKLSFKDAAVNRVMKQMAPAILGV	240
	orf20ng	YFDPVPTALAWAVFVGILQLGFLPWLAKLGLKLPKLNFKDAAVNRVMKQMAPAILGV	240
55	orf20.pep	SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT	300
	orf20ng	SVAQISLVINTIFASYLQSGSVSWMYADRMELPGGVLGAALGTILLPTLSKHSANQDT	300
60	orf20.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHAIAYSFG	360
	orf20ng	EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG	360
65	orf20.pep	LIGLIMIKVLAPGFYARQNIIXPVKIAIFTLICXQLMNLXFXGPLXXIGLSLAIGLGACI	420
	orf20ng	LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICXQLMNLAFIGPLKHAGLSLAIGLGACI	420
70	orf20.pep	NAGLLFYLLRRHGIYQXPQGLGSVLXQKCCSRSP	454
	orf20ng	NAGLLFFLFRKHGIYRPGQGLGQPSWRKCCSRSP	454

An ORF20ng nucleotide sequence <SEQ ID 119> was predicted to encode a protein having amino acid sequence <SEQ ID 120>:

```

1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL
51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA
101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVGS
151 LNSYHKFGIP AFTPTFLNIS FIVEALFFVP YFDPVPTALA WAVFVGILQ
5 LGFQLPWLAK LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN
201 TIFASYLQSG SVSWMYADR MMELPGGVLG AALGTILLPT LSKHSANQDT
251 EQFSALLDWG LRLCMLLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT
301 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNL
351 FIGPLKHAGL SLAIGLGACI NAGLLFFLLR KHGIYRPGQG LGQPSWRKCC
10 FIGPLKHAGL SLAIGLGACI NAGLLFFLLR KHGIYRPGQG LGQPSWRKCC
401 SRSF*
451

```

Further DNA sequence analysis revealed the following DNA sequence <SEQ ID 121>:

```

1 ATGAATATGC TTGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTG
51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTGCGCG
101 CCGGTATGGC GACGGATGCG TTTTGTGTCG CGTTCAAACG GCCCAACCTG
151 CTTCGCGCGC TGTTCGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCGCAT
201 TTTGGCGGAA TATAAGGAAA CGCGTTCTAA AGAGGCGAag gAGGCTTTTA
251 TCCGCCACGt tgcgggAatg CTGTCGTTTG TGCTGATcgt cGttacCGCG
301 CTGGGCATAC TTGCCGCGcc tTGGGTGATT TATGTTtccg CgcccGGCTT
351 TACCAAAGAC GCGGACAAGT TCCAACTTTC CATCAGCCTG CTGCGGATTA
20 CGTTTCCTTA TATATTATTG ATTTCTTGT CTCTTTTGT CGGCTCGATA
401 CTCAATTCCCT ACCATAAGTT CGGCATTCCC GCGTTTACG CCACGTTTTT
451 AAACATCTCT TTTATCGTAT TCGCACTGTT TTTCGTGCCG TATTTTCGATC
501 CGCCCGTTAC CGCGCTGGCG TGGGCGGTTT TTGTCGCGCG TATTTTGCAG
551 CTCGGTTTCC AACTGCCGTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC
601 CAAACTGAAT TTCAAAGATG CGGCGGTCAA CCGCGTCATG AAACAGATGG
25 701 CGCCTGCGAT TTTGGGCGTG agcgTGGCGC AAATTTCTTT GgttATCAAC
751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTatta
801 cgCCGACCGC ATGATGGAGc tgcgcCGGGG CGTGCTGGGG GCTGCACTCG
851 GTACAATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
30 901 GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCCTGT GCATGCTGCT
951 CGAGCTGCCG GCGGCGGcgg GACTGGCGGT ATTGTCGTTT CCGCTGGTGG
1001 CGAGCTGTTT TATGTACCGA GAATTCACGC TGTTTGACGC ACAAATGACG
1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGTT TAATTATGAT
1101 TAAAGTGTG GCATCCGCTT TTATGCGCG GCAAAACATC AAAACGCCCC
35 1151 TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTCGCC
1201 TTTATCGGTC CGTTGAAACA CGCCGGGCTT TCGCTCGCCA TCGGCTGGG
1251 CGCGTGCATC AACGCCGAT TGTGTTCTT CCTGTTGCGC AAACACGGTA
1301 TTTACCGGCC cggcaggggt tgggcggcgt TCTTGGCGAA AATGCTGCTC
1351 GCGCTCGCCG TGATGTGCGG CGGACTGTGG GCGGCGCAGG CTTGCTGCCC
40 1401 GTTCGAATGG GCGCACGCCG GCGGAATGCG GAAAGCGGGG CAGCTCTGCA
1451 TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCTCT GCGGCTTTG
1501 GGCTTCCGTC CGCGCCATT CAAACGCGTG GAAAGCTGA

```

This encodes the following amino acid sequence <SEQ ID 122; ORF20ng-1>:

```

1 MNMLGALAKV GSLTMVSRVL GFVRDVIAR AFGAGMATDA FFVAFKLPNL
45 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA
101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVGS
151 LNSYHKFGIP AFTPTFLNIS FIVEALFFVP YFDPVPTALA WAVFVGILQ
201 LGFQLPWLAK LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN
251 TIFASYLQSG SVSWMYADR MMELRRGVLG AALGTILLPT LSKHSANQDT
50 301 EQFSALLDWG LRLCMLLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT
351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNL
401 FIGPLKHAGL SLAIGLGACI NAGLLFFLLR KHGIYRPRG WAAFLAKMLL
451 ALAVMCGGLW AAQACLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAA
501 GFRPRHFKRV ES*

```

55 ORF20ng-1 and ORF20-1 show 95.7% identity in 512 aa overlap:

```

10 20 30 40 50 60
orf20-1.pep MNMLGALAKVGSGLTMVSRVLGFVRDVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
60 orf20ng-1 MNMLGALAKVGSGLTMVSRVLGFVRDVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
10 20 30 40 50 60
70 80 90 100 110 120
orf20-1.pep AQAFVPILAEYKETRSKEAAEFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPGFAQD
65 orf20ng-1 AQAFVPILAEYKETRSKEATEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPGFTKD

```

-121-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf20-1.pep	ADKFQLSIDLLRITFPYILLISLSSSVGSLNSYHKFGIPAFTPTFLNVSFIVFALFFVP					
	orf20ng-1	ADKFQLSISLLRITFPYILLISLSSSVGSLNSYHKFGIPAFTPTFLNVSFIVFALFFVP					
		130	140	150	160	170	180
10	orf20-1.pep	YFDPPTALAWAVFVGILQLGFQLPWLAKLGLKLPKLSFKDAAVNRVMQMAPPAILGV					
	orf20ng-1	YFDPPTALAWAVFVGILQLGFQLPWLAKLGLKLPKLNFKDAAVNRVMQMAPPAILGV					
		190	200	210	220	230	240
15	orf20-1.pep	SVAQVSLVINTIFASYLQSGSVSWMYADRMELPSGVLGAALGTILLPTLSKHSANQDT					
	orf20ng-1	SVAQISLVINTIFASYLQSGSVSWMYADRMELRRGVLGAAALGTILLPTLSKHSANQDT					
		250	260	270	280	290	300
20	orf20-1.pep	EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG					
	orf20ng-1	EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG					
25		310	320	330	340	350	360
	orf20-1.pep	LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHHVGLSLAIGLGACI					
	orf20ng-1	LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI					
30		370	380	390	400	410	420
	orf20-1.pep	NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRKAG					
	orf20ng-1	NAGLLFFLLRKHGIYRPGRGWAAFLAKMLLALAVMCGGLWAAQACLPFEWAHAGGMRKAG					
35		430	440	450	460	470	480
	orf20-1.pep	QLCILIAGGGGLYFASLAALGFRPRHFKRVENX					
	orf20ng-1	QLCILIAGGGGLYFASLAALGFRPRHFKRVESX					
40		490	500	510			

In addition, ORF20ng-1 shows significant homology with a virulence factor of *S. typhimurium*:

45	sp P37169 MVIN SALTY VIRULENCE FACTOR MVIN pir S40271 mviN protein - Salmonella typhimurium gi 438252 (Z26133) mviB gene product [Salmonella typhimurium] gnl PID d1005521 (D25292) ORF2 [Salmonella typhimurium] Length = 524 Score = 1573 (750.1 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220 Identities = 309/467 (66%), Positives = 368/467 (78%)
50	Query: 1 MNMLGALAKVGS LTMVSRVLGFRDVTIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 60 MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDAFFVAFKLPNLLRR+FAEGAF Sbjct: 14 MNLLKSLAAVSSMTMFSRVLGFAIDAIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF 73
55	Query: 61 AQAFVPILAEYKETSKEATEAFIRHVAGMLS FVLIVVTALGILAAPWVIYVSAPGFTKD 120 +QAFVPILAEYK + +EAT F+ +V+G+L+ L VVT G+LAAPWVI V+APGF Sbjct: 74 SQAFVPILAEYKSKQGEATRI FVAYVSGLLTALAVVTVAGMLAAPWVIMVTAPGFADT 133
60	Query: 121 ADKFQLSISLLRITFPYILLISLSSSVGSLNSYHKFGIPAFTPTFLNISFIVFALFFVP 180 ADKF L+ LLRITFPYILLISL+S VG+ILN++++F IPAF PTFLNIS I FALF P Sbjct: 134 ADKFALTQLLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISMIGFALFAAP 193
65	Query: 181 YFDPPTALAWAVFVGILQLGFQLPWLAKLGLKLPKLNFKDAAVNRVMQMAPPAILGV 240 YF+PPV ALAWAV VGG+LQL +QLP+L K+G L LP++NF+D RV+KQM PAILGV Sbjct: 194 YFNPPVLALAWAVTVGGVLQLVYQLPYLKKIGMLVLPRIINFRDGTAMRVVQMGPAILGV 253
70	Query: 241 SVAQISLVINTIFASYLQSGSVSWMYADRMELRRGVLGAAALGTILLPTLSKHSANQDT 300 SV+QISL+INTIFAS+L SGSVSWMYADR+ME GVLG ALGTILLP+LSK A+ + Sbjct: 254 SVSQISLIINTIFASFLASGSVSWMYADRLMEFPSGVLGVALGTILLPSLSKSFASGNH 313

-122-

Query: 301 EQFSALLDWGLRLCMLLTPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHAIAYSFG 360
 +++ L+DWGLRLC LL LP+A L +L+ PL +LF Y +FT FDA MTQ ALIAYS G
 Sbjct: 314 DEYCRLMDWGLRLCFLALPSAVALGILAKPLTVSLFQYQKFTAFDAAMTORALIAYSFG 373

5 Query: 361 LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI 420
 LIGLI++KVLA GFY+RQ+IKTPVKIAI TLI TQLMNLAFIGPLKHAGLSL+IGL AC+
 Sbjct: 374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

10 Query: 421 NAGLLFFLLRKHGIIYRPGRWXXXXXXXXXXXXXVMCGGLWAAQACLP 467
 NA LL++ LRK I+ P GW VM L+ +P
 Sbjct: 434 NASLLYWQLRKQNIFTPQPGWWMFLMRIISVLVMAAVLFGVLHIMP 480

Score = 70 (33.4 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220
 Identities = 14/41 (34%), Positives = 23/41 (56%)

15 Query: 469 EWAHAGGMRKAGQLCILIIVGGGLYFASLAALGFRPRHFKR 509
 EW+ + + +L ++ G YFA+LA LGF+ + F R
 Sbjct: 481 EWSQGSMLWRLRLMAVVIAGIAAYFAALAVLGFKVKEFVR 521

- 20 Based on this analysis, including the homology with a virulence factor from *S.typhimurium*, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 15

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 123>:

25 1 atGATTAAAA TCAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGCCGTT TACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGCGG
 101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
 151 GTCAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGCT
 201 GTTTACTGCG CCGGCTTCAG GcAAAATCGC CGCGATTAC CGTGGCGAAA
 30 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGCAA CGACGAAATC
 301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACCTAA GCGGCGAAGA
 351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
 401 GTCCGTTGAG CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
 451 GTCAATGCGA tGGACACCAA TCCG..

- 35 This corresponds to the amino acid sequence <SEQ ID 124; ORF22>:

1 MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA
 51 VKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEXNDEI
 101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
 151 VNAMDTNP..

- 40 Further work revealed the complete nucleotide sequence <SEQ ID 125>:

1 ATGATTAAAA TCAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGCCGTT TACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGCGG
 101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
 151 GTCAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGCT
 45 201 GTTTACTGCG CCGGCTTCAG GcAAAATCGC CGCGATTAC CGTGGCGAAA
 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGCAA CGACGAAATC
 301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACCTAA GCGGCGAAGA
 351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
 401 GTCCGTTGAG CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
 50 451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTCAATTAT
 501 CAAAGAAGCC GCCGAGGATT TCAAACGCGG CCTGTTGGTA TTGAGCCGTT
 551 TGACCGAAGC CAAAATCCAT GTTTGTAAGG CAGCTGGCGC AGACGTGCCG
 601 TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC
 651 TGCCGGTTTG AGTGCGCAGC ACATTCAATT CATCGAGCCG GTCGCGCGGA
 701 ATAAAACCGT GTGGACCATC AATTATCAAG ATGTAATTAC CATTGGCCGT
 751 TTGTTTGCAA CAGGCGGTCT GAACACCGAG CGCGTGATTG CCCTAGGTGG
 801 TTCTCAAGTC AACAAACCGC GCCTCTTGCG TACCGTTTGG GGTGCGAAAG
 851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACACAGACAA CCGCGTGATT
 901 TCCGGTTCGG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT

5 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
 1001 AGCTGTTTCGG CTGGGTTGCG CCGCAGCCGG ACAATACTC CATCACGCGT
 1051 ACAACCCTCG GCCATTTCCT GAAAAACAAA CTCTTCAAGT TCAACACAGC
 1101 CGTCAACGGC GGCAGCCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
 1151 TGATGCCCTT GGATATCCTG CCCACCCTGC TTTTGCGCGA TTTAATCGTC
 1201 GGCATACCG ACAGCGCGCA GGCATTGGGT TGCTTGAAT TGGACGAAGA
 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC
 1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA

This corresponds to the amino acid sequence <SEQ ID 126; ORF22-1>:

10 1 MIKIKKGLNL PIAGRPEQAV YDGPATEVA LLGEEYAGMR PSMKVKEGDA
 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEGNDEI
 101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
 151 VNAMDTNPLA ADPTVVIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADV
 201 SENAANIETH EFGGPHFAGL SGTHIHFIET VGANKTVWTI NYQDVITIGR
 15 251 LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDTDNRVI
 301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELFGWVA PQPDKYSITR
 351 TTLGHFLKNK LFKFNTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV
 401 GDTSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG*

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 127>:

20 1 ATGATTAAAA TCAAAAAGG TCTAACCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGTCATT TATGACGGG CCGTCATTAC CGAAGTCGCG TTGCTTGGCG
 101 AAGAATATGC CGGTATGCGC CCTNGATGA AAGTCAAGGA AGGCGATGCC
 151 GTCAAAAAG GCGAAGTGCT GTTTGAAGAC AAAAAGNATC CGGGCGTGGT
 201 GTTACCGCG CCNGTTTCAG GCAAAATCGC CGCCATCCAT CGCGGCGAAA
 25 251 AGCGCGTACT TCAGTCGCTG GTGATTGCGG TTGAAGGCAA CGACGAAATC
 301 GAGTTCGAAC GCTACGCGCC CGAAGCGTTG GCAAACTTAA CGGGCGANGA
 351 ANTNNNGNGC AATCTGATCC AATCCGGTTT GTGGACTGCG CTGCGTANCC
 401 GTCCGTTTCA CAAAATCCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
 451 GTCAATGCGA TGGACACCAA TCCGCTNGCG GCAGACCCCTG TGGTTGTGAT
 30 501 CAAAGAAGCC GNCGANGATT TCAGACGANG TNGTCTGTA TTGAGCCGTT
 551 TGACCGAGCG TAAATCCAT GTGTGTAAGG CAGCTGGCGC AGACGTGCCG
 601 TCTGAAAATG CTGCCAATC CGAAACACAT GAATTCGCGG GCCCGCATCC
 651 GGCCGGTTTG AGTGGCAGC ACATTTCATT CATTTAGCCG GTCGGTGCAA
 701 ACAAACCGT TTGGACCATC AATTATCAAG ATGTAATTGC CATCGGACGT
 35 751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CTTTGGGTGG
 801 TTCTCAAGTC AACAAACAC GCCTCTTGCG TACCGTTTGG GGTGCGAAAG
 851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACGCAGACAA CCGCGTGATT
 901 TCCGTTTCGG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT
 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
 40 1001 AGCTGTTTCGG CTGGGTTGCG CCGCAGCCGG ACAATACTC CATCACGCGT
 1051 ACGACCCTCG GCCATTTCCT GAAAAACAAA CTCTTCAAGT TCACGACAGC
 1101 CGTCAACGGT GCGCAGCCGG CCATGGTGCC GATTGGTACT TACGAGCGCG
 1151 TAATGCCGCT AGACATCCTG CCTACCCTGC TTTTGCGCGA TTTAATCGTC
 1201 GGCATACCG ACAGCGCGCA AGCATTGGGT TGCTTGAAT TGGACGAAGA
 45 1251 AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATANGGCC
 1301 CGCTGTTGCG TAAGTGCTG GAAACCNTTG AGAAGGAAGG CTGA

This encodes a protein having amino acid sequence <SEQ ID 128; ORF22a>:

50 1 MIKIKKGLNL PIAGRPEQVI YDGPVITEVA LLGEEYAGMR PSMKVKEGDA
 51 VKKGQVLFED KKNPGVVFTA PVSGKIAAIH RGEKRVLSV VIAVEGNDEI
 101 EFERYAPEAL ANLSGXEXX NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
 151 VNAMDTNPLA ADPVVVIKEA XXDFRXXLV LSRLTERKIH VCKAAGADV
 201 SENAANIETH EFGGPHFAGL SGTHIHFIET VGANKTVWTI NYQDVIAIGR
 251 LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDADNRVI
 301 SGSVLNGAIT QGAHDYLGRI HNQISVIEEG RSKELFGWVA PQPDKYSITR
 55 351 TTLGHFLKNK LFKFNTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV
 401 GDTSAQALG CLELDEEDLA LCSFVCPGKY EXGPLLRKVL ETXEKEG*

The originally-identified partial strain B sequence (ORF22) shows 94.2% identity over a 158aa overlap with ORF22a:

		10	20	30	40	50	60
60	orf22.pep	MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKKGQVLFED					
	orf22a	MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPSMKVKEGDAVKKKGQVLFED					

-124-

		10	20	30	40	50	60
5	orf22.pep	70	80	90	100	110	120
	orf22a	70	80	90	100	110	120
10	orf22.pep	130	140	150			
	orf22a	130	140	150	160	170	180
The complete strain B sequence (ORF22-1) and ORF22a show 94.9% identity in 447 aa overlap:							
15	orf22a.pep	10	20	30	40	50	60
	orf22-1	10	20	30	40	50	60
20	orf22a.pep	70	80	90	100	110	120
	orf22-1	70	80	90	100	110	120
25	orf22a.pep	130	140	150	160	170	180
	orf22-1	130	140	150	160	170	180
30	orf22a.pep	190	200	210	220	230	240
	orf22-1	190	200	210	220	230	240
35	orf22a.pep	250	260	270	280	290	300
	orf22-1	250	260	270	280	290	300
40	orf22a.pep	310	320	330	340	350	360
	orf22-1	310	320	330	340	350	360
45	orf22a.pep	370	380	390	400	410	420
	orf22-1	370	380	390	400	410	420
50	orf22a.pep	430	440				
	orf22-1	430	440				

Further work identified a partial gene sequence <SEQ ID 129> from *N.gonorrhoeae*, which encodes the following amino acid sequence <SEQ ID 130; ORF22ng>:

65	1	MIKIKKGLNL	PIAGRPEQVI	YDGAITEVA	LLGEEYVGM	PSMKIKEGEA
	51	VKKGQVLFED	KKNPGVVFTA	PASGKIAAIH	RGEKRVLSV	VIAVEGNDEI
	101	EFERYVPEAL	AKLSSEKVR	NLIQSGLWTA	LRTRPFSKIP	AVDAEPFAIF

151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADV
 201 SENAANIETH EFGGPHAGL SGTHIHFIET VGANKTVWTI NYQDVIAIGR
 251 LFTVTRLNTE RVVALGGLQV NKPRLLRTVL GAKVSQLTAG ELVDADNRVI
 301 SGSVLNGAIA QGAHDYLGRY HN*

5 Further work identified complete gonococcal gene <SEQ ID 131>:

1 ATGATTAAAA TCAAAAAAGG TCTAAATCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGTCATT TATGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG
 101 AAGAATATGT CGGCATGCGC CCCTCGATGA AAATCAAGGA AGGTGAAGCC
 151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAGAATC CGGGCGTAGT
 201 ATTTACTGCG CCGGCTTCAG GCAAAATCGC CGCTATTAC CGTGGCGAAA
 251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301 GAGTTCGAAC GCTACGTACC TGAAGCGCTG GCAAAATTGA GCAGCGAAAA
 351 AGTGGCGCGC AACCTGATT CATTAGGCTT ATGGACTGCG CTTCCGACCC
 401 GTCCGTTTCAG CAAAATCCCT GCCGTAGATG CCGAGCCGTT CGCCATCTTC
 15 451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCTA CGGTCAATCAT
 501 CAAAGAAGCC GCCGAAGACT TCAAACGCGG CCTGTTGGTA TTGAGCCGCC
 551 TGACCGAAGC TAAATCCAT GTGTGTAAAG CAGCAGGCGC AGACGTGCCG
 601 TCTGAAATG CTGCCAATAT CGAAACACAT GAATTTGGCG GCCCGCATCC
 651 TGCCGCTTGG AGTGGCACGC ACATTCAATT CATCGAGCCA GTCGGCGCGA
 20 701 ATAAACCGT GTGGACCATC AATTATCAAG AGTGATGTC TATCGGACGT
 751 TTGTTCTGTA CAGGCCGTCT GAATACCGAG CGCGTGGTTG CCTTGGGCGG
 801 CCTGCAAGTC AACAAACGCG GCCTCTTGCG TACCGTTTG GGTGCGAAGG
 851 TGTCTCAACT TACCGCCGCG GAATTGGTTG ACGCGGACAA CCGCGTGATT
 901 TCCGGTTCGG TATTGAACGG TCGATTGCA CAAGGCGCGC ATGATTATTT
 25 951 GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
 1001 AGCTGTTGCG CTGGGTGCG CCGCAGCCGG ACAAACTACT CATCACGCGC
 1051 ACCACTCTCG GCCATTTCTT AAAAAACAAA CTCTTCAAGT TCACGACAGC
 1101 CGTCAACGGC GCGGACCGCG CCATGGTACC GATCGGCACT TATGACGCGC
 1151 TAATGCCGTT GGACATCTCG CCTACCTTGC TTTGCGCGA TTTAATCGTC
 30 1201 GGCGATACCG ACAGCGCGCA GGCTTTGGGT TGCTTGAAT TGGACGAGA
 1251 AGACCTCGCT TGTGTCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC
 1301 CGCTGTTGCG CAAAGTGCTG GAAACCATTT AGAAGGAAGG CTGA

This encodes a protein having amino acid sequence <SEQ ID 132; ORF22ng-1>:

35 1 MIKIKKGLNL PIAGRPEQVI YDGPATEVA LLGEEYVGM PSMKIKEGEA
 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLSV VIAVEGNDEI
 101 EFERYVPEAL AKLSSEKVR NLIQSLWTA LRTRPFSKIP AVDAEPFAIF
 151 VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADV
 201 SENAANIETH EFGGPHAGL SGTHIHFIET VGANKTVWTI NYQDVIAIGR
 40 251 LFTVTRLNTE RVVALGGLQV NKPRLLRTVL GAKVSQLTAG ELVDADNRVI
 301 SGSVLNGAIA QGAHDYLGRY HNQISVIEEG RSKELFGWVA PQPDKYSITR
 351 TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV
 401 GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG*

The originally-identified partial strain B sequence (ORF22) shows 93.7% identity over a 158aa

45 overlap with ORF22ng:

orf22.pep MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED 60
 orf22ng MIKIKKGLNLPIAGRPEQVIYDGPATEVALLGEEYVGMPSMKIKEGEAVKKGQVLFED 60
 50 orf22.pep KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGEEVRR 120
 orf22ng KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYVPEALAKLSSEKVR 120
 orf22.pep NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP 158
 55 orf22ng NLIQSLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV 180

The complete sequences from strain B (ORF22-1) and gonococcus (ORF22ng) show 96.2% identity in 447 aa overlap:

60 orf22-1.pep MIKIKKGLNLPIAGRPEQAVYDGPATEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED

-126-

	or22ng-1	MIKIKKGLNLP 10	IAGRPEQVIYD 20	GPAITEVALLG 30	EYVGMPSMKI 40	KEGEAVKKGQ 50	VLFED 60
5	or22-1.pep	70	80	90	100	110	120
	or22ng-1	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI	AVEGNDEIEFERYAPEALANLSGEEVRR				
10	or22-1.pep	130	140	150	160	170	180
	or22ng-1	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFV	NAMDTNPLAADPTVIIKEAAEDFKRGLLV				
15	or22-1.pep	190	200	210	220	230	240
	or22ng-1	LSRLTERKIHVCKAAGADVPSENAANIETHE	FGGPHPAGLSGTHIHFIIEPVGANKTVWTI				
20	or22-1.pep	250	260	270	280	290	300
	or22ng-1	NYQDVITIGRLFATGRNLNTERVIALGGSQV	NKPRLRLTVLGAKVSQITAGELVDTDNRVI				
25	or22-1.pep	310	320	330	340	350	360
	or22ng-1	SGSVLNGAITQGAHDYLGRYHNQISVIEEGR	SKELFGWVAPQDPKYSITRTTLGHFLKNK				
30	or22-1.pep	370	380	390	400	410	420
	or22ng-1	LKFNTAVNGGDRAMVPIGTYERVMPLDILPT	LLLRDLIVGDTSAQALGCLELDEEDLA				
35	or22-1.pep	430	440				
	or22ng-1	LCSFVCPGKYEYGPLLRRKVLETIEKEGX					
40	or22-1.pep	430	440				
	or22ng-1	LCSFVCPGKYEYGPLLRRKVLETIEKEGX					

Computer analysis of these sequences gave the following results:

Homology with 48kDa outer membrane protein of *Actinobacillus pleuropneumoniae* (accession number U24492).

ORF22 and this 48kDa protein show 72% aa identity in 158aa overlap:

50	Orf22	1	MIKIKKGLNLP MI IKKGL+LPIAG P Q +++G + EVA+LGEEY GMRPSMKV+EGD VKKGQVLFED	60
	48kDa	1	MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMPSMKVREGDVVKKGQVLFED	60
55	orf22	61	KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI KKNPGVVFTAPASG + I+RGEKRVLSVVI VE +++I F RY LA+LS E+v++	120
	48kDa	61	KKNPGVVFTAPASGTVVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ	120
60	orf22	121	NLIQSGLWTALRTRPFSKIPAVDAEPFAIFV NLI+SGLWTA RTRPFSK+PA+DA P +IFVNAMDTNP	158
	48kDa	121	NLIESGLWTAFTTRPFSKVPALDAIPSSIFVNAMDTNP	158

ORF22a also shows homology to the 48kDa *Actinobacillus pleuropneumoniae* protein:

gi|1185395 (U24492) 48 kDa outer membrane protein (*Actinobacillus pleuropneumoniae*)
Length = 449

65 Score = 530 bits (1351), Expect = e-150

-127-

Identities = 274/450 (60%), Positives = 323/450 (70%), Gaps = 4/450 (0%)

Query: 1 MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMVKVEGDVKKGQVLFED 60
 5 Sbjct: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGM RPSMKVREGDVVKKGQVLFED 60

Query: 61 KKKPGVVFTAPVSGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYAPEALANLSGXEXX 120
 10 Sbjct: 61 KKNPGVVFTAPASGTVVTINRGEKRVLSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

Query: 121 NLIQSGLWLTALRXRPFPSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRXXLV 180
 Sbjct: 121 NLIESGLWTAFRTRPFPSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGTLV 180

Query: 181 LSRL--TERKIHVCKAAGADVP--SENAANIETHEFGGPHAGLSGTHIHFIIEPVGANKTV 237
 15 Sbjct: 181 LTRLFNGQKPVYLCCKDADSNIPSPAIEGITIKSFSGVHPAGLVGTHIHFDVPVGATKQV 240

Query: 238 WTINYQDVIAIGRLFATGRNTERVIALGGSQVNKPRLRLTVLGAKVSQITAGELVDADN 297
 20 Sbjct: 241 WHLNYQDVIAIGKLFITGELFTDRIISLAGPQVKNPRLVTRRLGANLSQLTANELNAGEN 300

Query: 298 RVISGSLNGAITQGAHDYLGRIYHNQISVIEEGRSKELFGWVAPQPDKYSITRTTLGHFL 357
 25 Sbjct: 301 RVISGSLVSGATAAGPVDYLGRIYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

Query: 358 KNKLFKFTTAVNGGDRAMVPIGTIERVMXXXXXXXXXXXXXXXXXVGDTSAQXXXXXXXXXX 417
 Sbjct: 361 K-KLFNFTTAVHGGGERAMVPIGAYERVMPLDIPTLLRLDLAAGDTSQNLGCLELDEE 419

Query: 418 XXXXXSFVCPGKYEXGPLLRKVLETXEKEG 447
 30 ++VCPGK GP+LR LE EKEG

ORF22ng-1 also shows homology with the OMP from *A. pleuropneumoniae*:

gil1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus
 35 pleuropneumoniae] Length = 449
 Score = 555 bits (1414), Expect = e-157
 Identities = 284/450 (63%), Positives = 337/450 (74%), Gaps = 4/450 (0%)

Query: 27 MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYVGM RPSMKIKEGEAVKKGQVLFED 86
 40 Sbjct: 1 MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGM RPSMKVREGDVVKKGQVLFED 60

Query: 87 KKNPGVVFTAPASGKIAAIHRGEKRVLSVVI AVEGNDEIEFERYVPEALAKLSSEKVR 146
 45 Sbjct: 61 KKNPGVVFTAPASG + I+RGEKRVLSVVI VEG+++I F RY LA LS+E+V++ 120

Query: 147 NLIQSGLWLTALRTRPFPSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV 206
 Sbjct: 121 NLIESGLWTAFRTRPFPSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGTLV 180

Query: 207 LSRL--TERKIHVCKAAGADVP--SENAANIETHEFGGPHAGLSGTHIHFIIEPVGANKTV 263
 Sbjct: 181 LTRLFNGQKPVYLCCKDADSNIPSPAIEGITIKSFSGVHPAGLVGTHIHFDVPVGATKQV 240

Query: 264 WTINYQDVIAIGRLFVTRNTERVIALGGLQVNKPRLRLTVLGAKVSQITAGELVDADN 323
 55 Sbjct: 241 WHLNYQDVIAIGKLFITGELFTDRIISLAGPQVKNPRLVTRRLGANLSQLTANELNAGEN 300

Query: 324 RVISGSLNGAIAQGAHDYLGRIYHNQISVIEEGRSKELFGWVAPQPDKYSITRTTLGHFL 383
 60 Sbjct: 301 RVISGSLVSGATAAGPVDYLGRIYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

Query: 384 KNKLFKFTTAVNGGDRAMVPIGTIERVMXXXXXXXXXXXXXXXXXVGDTSAQXXXXXXXXXX 443
 65 Sbjct: 361 K-KLFNFTTAVHGGGERAMVPIGAYERVMPLDIPTLLRLDLAAGDTSQNLGCLELDEE 419

Query: 444 XXXXXSFVCPGKYEYGPLLRKVLETIEKEG 473
 ++VCPGK YGP+LR LE IEKEG
 70 Sbjct: 420 DLALCTYVCPGKNYGPMLRAALEKIEKEG 449

Based on this analysis, including the homology with the outer membrane protein of *Actinobacillus pleuropneumoniae*, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF22-1 (35.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 5A shows the results of affinity purification of the GST-fusion protein, and Figure 5B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 5C). These experiments confirm that ORF22-1 is a surface-exposed protein, and that it is a useful immunogen.

10 Example 16

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 133>:

```

1   .GCGnCGnAAA TCATCCATCC CC..nACGTC GTAGGCCCTG AAGCCAAC TG
51  GTTTTATATG GTAGCCAGTA CGTTTGTGAT TGCTTTGATT GGTATTATTTG
101 TTACTGAAAA AATCGTCGAA CCGCAATTGG GCCCTTATCA ATCAGATTTG
15  TCACAAGAAG AAAAAGACAT TCGGCATTCC AATGAAATCA CGCCTTTGGA
201 ATATAAAGGA TTAATTTGGG CTGGCGTGGT GTTGTGTGCC TTATCGCCCC
251 TATTGGCTTG GAGCATCGTC CCTGCCGACG GTATTTTGCG TCATCCTGAA
301 ACAGGATTGG TTTCCGGTTC GCCGTTTTTA AAATCGATTG TTGTTTTTAT
351 TTTCTTGTG TTTGCACTGC CGGGCATTGT TTATGGCCGG GTAACCCGAA
20  GTTTGCGCGG CGAACAGGAA GTCGTTAATG CGmyGGCCGA ATCGATGAGT
451 ACTCTGGsGC TTTmTTTGSW CAkCATCTTT TTTGCCGCAC AGTTTGTGCG
501 ATTTTTTAAT TGGACGAATA TTGGGCAATA TATTGCCGTT AAAGGGGCGA
551 CGTTCTTAAA AGAAGTCGGC TTGGGCGGCA GCGTGTGTGT TATCGGTTTT
601 ATTTTAATTT GTGCTTTTAT CAATCTGATG ATAGGCTCCG CCTCCGCGCA
25  ATGGGCGGTA ACTGCCCGCA TTTTCGTCCC TATGCTGATG TTGGCCGGCT
701 ACGCGCCCGA AGTCATTCAA GCCGCTTACC GCATCGGTGA TTCCGTTTACC
751 AATATTATTA CGCCGATGAT GAGTTATTTT GGGCTGATTA TGGCGACGGT
801 GrkCmmnTAC AAAAAGATG CGGGCGTGGG TaCGcTGATT wCTATGATGT
851 TGCCGTATTC CGCTTCTTC TTGATTGCgT GGATTGCCTT ATTCTGCATT
30  TGGGTATTTg TTTTGGGCCT GCCCGTCGGT CCCGCGCGGC CCACATTCTA
951 TCCCGCACCT TAA

```

This corresponds to the amino acid sequence <SEQ ID 134; ORF12>:

```

1   .AXXIHPXXV VGPEANWFFM VASTFVIALI GYFVTEKIVE PQLGPYQSDL
35  SQEEKDIRHS NEITPLEYKG LIWAGVVFVA LSALLAWSIV PADGILRHPE
101 TGLVSGSPFL KSIVVFIFLL FALPGIVYGR VTRSLRGEQE VVNAXAESMS
151 TLXLXLXIF FAAQFVAFEN WTNIGQYIAV KGATFLKEVG LGGSVLFIGF
201 ILICAFINLM IGSASAQWAV TAPIFVPLM LAGYAPEVIQ AAYRIGDSVT
251 NIITPMMSYF GLIMATVXXY KKDAGVGTLI XMMLPYSAFF LIAWIALFCI
301 WVFVLGLPVG PGAPTFYPAP *

```

40 Further sequence analysis revealed the complete DNA sequence <SEQ ID 135> to be:

```

1   ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTAC GCACAGTCGA
51  ATGGCTGGGC AATATGTTGC CGCATCCGGT TACGCTTTT ATTATTTC A
101 TTGTGTTATT GCTGATTGCC TCTGCCGTCG GTGCGTATTT CGGACTATCC
151 GTCCCGGATC CGCGCCCTGT TGGTGCAGAA GGACGTGCCG ATGACGGTTT
45  GATTTACATT GTCAGCCTGC TCAATGCCGA CGGTTTTATC AAAATCCTGA
201 CGCATACCGT TAAAAATTTT ACCGGTTTCG CGCCGTTGGG AACGGTGTG
251 GTTTCTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCCGC
301 ATTAATGCGC TTATTGCTCA CAAATCGCC ACGCAAAC TC ACTACTTTA
351 TGGTTGTTT TACAGGGATT TTATCTAATA CCGCTTCTGA ATTGGGCTAT
40  TCGCTCCTAA TCCCTTTGTC CGCCATCATC TTTATTCCC TCGGCCGCCA
451 TCCGCTTGCC GGTCTGGCTG CGGCTTTCGC CGGCTTTCG GCGGTTTATT
501

```

551 CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTTGGC AGGCATCACC
 601 CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
 651 CAACTGGTTT TTTATGGTAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
 701 ATTTTGTAC TGAATAAATC GTCGAACCGC AATTGGGCCC TTATCAATCA
 751 GATTGTGCAC AAGAAGAAAA AGACATTCGG CATTCCAATG AAATCACGCC
 801 TTTGGAATAT AAAGGATTAA TTTGGGCTGG CGTGGTGTTC GTTGCCATTAT
 851 CCGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
 901 CCTGAAACAG GATTGGTTTC CGGTCGCGG TTTTAAATAT CGATTGTTGT
 951 TTTTATTTTC TTGTTGTTTG CACTGCCGGG CATTTGTTTAT GGCCGGGTAA
 1001 CCCGAAGTTT GCGCGGCGAA CAGGAAGTCG TTAATGCGAT GGCCGAATCG
 1051 ATGAGTACTC TGGGGCTTTA TTTGGTCATC ATCTTTTTTG CCGCACAGTT
 1101 TGTGCGATTT TTTAATTGGA CGAATATTGG GCAATATATT GCCGTAAAG
 1151 GGGCGACGTT CTAAAAGAA GTCGGCTTGG GCGGCAGCGT GTTGTATTATC
 1201 GGTTTTATTT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCCTC
 1251 CGCGCAATGG GCGGTAAGTG CGCGGATTTT CGTCCCTATG CTGATTTGG
 1301 CCGGCTACGC GCCCGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
 1351 GTTACCAATA TTATTACGCC GATGATGAGT TATTTCGGGC TGATTATGGC
 1401 GACGGTGATC AAATACAAA AAGATGCGGG CGTGGGTACG CTGATTCTTA
 1451 TGATGTTGCC GTATTCCGCT TTCTTCTTGA TTGCGTGGAT TGCCTTATTC
 1501 TGCATTGGG TATTGTTTT GGGCCTGCCG GTCGTCGCCG GCGCGCCAC
 1551 ATTCTATCCC GCACCTTAA

This corresponds to the amino acid sequence <SEQ ID 136; ORF12-1>:

1 MSQTDTRDQ RFLRTVEWLG NMLPHVTLF IIFIVLLLIA SAVGAYFGLS
 51 VPDPRPVGAK GRADDGLIYI VSLNADGFI KILHTVKNF TGFAPLGTVL
 101 VSLLGVGIAE KSGSLISALMR LLLTKSPRKL TTFMVVFTGI LSNTASELGY
 151 VVLIPLSAII FHSILGRHPLA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT
 201 QQAQIIHPD YVVGPEANWF FMVASTFVIA LIGYFVTEKI VEPQLGPYQS
 251 DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
 301 PETGLVSGSP FLKSIVVFIF LFLALPGIVY GRVTRSLRGE QEVVNMAES
 351 MSTLGLYLV IFFAAQFVAF FWNITNIGQYI AVKGATFLKE VGLGGSVLFI
 401 GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGYAPEV IQAAYRIGDS
 451 VTNIIIPMMS YFGLIMATVI KYKDGAVGT LISMLPYSA FFLIAWIALF
 501 CIWVFLGLP VGPGAPTFYP AP*

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF12 shows 96.3% identity over a 320aa overlap with an ORF (ORF12a) from strain A of *N.*

meningitidis:

				10	20	30
40	orf12.pep			AXXIHPXXVVGPEANWFMVASTFVIALI		
	orf12a	AAAFAGVSGGYSANLFLGTIDPLLAGITQQAQIIHPDYVVGPEANWFMVASTFVIALI				
		180 190 200 210 220 230				
45	orf12.pep	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV	40 50 60 70 80 90			
	orf12a	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV				
		240 250 260 270 280 290				
50	orf12.pep	PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS	100 110 120 130 140 150			
	orf12a	PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNMAESMS				
		300 310 320 330 340 350				
55	orf12.pep	TLXLXLXIIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLM	160 170 180 190 200 210			
	orf12a	TLGLYLVIIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLM				
60		360 370 380 390 400 410				
	orf12.pep	IGSASAQWAVTAPIFVPMMLAGYAPEVIAQAYRIGDSVTNIIIPMMSYFGLIMATVXXY	220 230 240 250 260 270			

orf12a	IGSASAQWAVTAPIFVPMMLLAGYAPEVIQAAAYRIGDSVTNIITPMMSYFGLIMATVIKY
	420 430 440 450 460 470
5	orf12.pep
	280 290 300 310 320
	KKDAGVGT LIXMMLPYSAFFLI AWIALFCI WVFVLGLPVGPGAPT FYPAPX
	orf12a
	KKDAGVGT LISMMLPYSAFFLI AWIALFCI WVFVLGLPVGPGAPT FYPAPX
	480 490 500 510 520

The complete length ORF12a nucleotide sequence <SEQ ID 137> is:

10	1	ATGAGTCAAA	CCGATACGCA	ACGGGACGGA	CGATTTTAC	GCACAGTCGA
	51	ATGGCTGGGC	AATATGTTGC	CGCACCCGGT	TACGCTTTT	ATTATTTTCA
	101	TTGTGTTATT	GCTGATTGCC	TCTGCCCGCG	GTGCGTATT	CGGACTATCC
	151	GTCCCCGATC	CGCGCCCTGT	TGGTGCAGAA	GGACGTGCCG	ATGACGGTTT
	201	GATTCACGTT	GTCAGCCTGC	TCGATGCTGA	CGGTTTGATC	AAAATCCTGA
15	251	CGCATACCGT	TAAAATTTT	ACCGGTTTCG	CGCGTTGGG	AACGGTGTG
	301	GTTTCTTTAT	TGGGCGTGGG	GATTGCGGAA	AAATCGGGCT	TGATTTCCGC
	351	ATTAATGCGC	TTATTGCTCA	CAAAATCTCC	ACGCAAACTC	ACTACTTTTA
	401	TGGTTGTTTT	TACAGGGATT	TTATCTAATA	CCGCTTCTGA	ATTGGGCTAT
	451	GTCGTCCTAA	TCCCTTTGTC	CGCCATCATC	TTTCATTCCC	TCGGCCGCCA
20	501	TCCGCTTGCC	GGTCTGGCTG	CGGCTTTCGC	CGCGTTCGC	GGCGGTTATT
	551	CGGCCAATCT	GTTCTTAGGC	ACAATCGATC	CGCTCTGGC	AGGCATCACC
	601	CAACAGGCGG	CGCAAATCAT	CCATCCCGAC	TACGTCGTAG	GCCCTGAAGC
	651	CAACTGGTTT	TTTATGGTAG	CCAGTACGTT	TGTGATTGCT	TTGATTGGTT
	701	ATTTTGTAC	TGAAAAAATC	GTCGAACCGC	AATTGGGCCC	TTATCAATCA
25	751	GATTGTGAC	AAGAAGAAAA	AGACATTGCA	CATTCCAATG	AAATCAGGCC
	801	TTTGGAATAT	AAAGGATTAA	TTTGGGCTGG	CGTGGTGTG	GTTGCGTTAT
	851	CCGCCCTATT	GGCTTGGAGC	ATCGTCCCTG	CCGACGGTAT	TTTGCCTCAT
	901	CCTGAAACAG	GATTGGTTTC	CGGTTGCGCG	TTTTTAAAT	CAATTGTTGT
	951	TTTTATTTTC	TTGTTGTTTG	CACTGCCGGG	CATTGTTTAT	GGCCGGGTAA
30	1001	CCCGAAGTTT	GCGCGGCGAA	CAGGAAGTCG	TTAATGCGAT	GGCCGAATCG
	1051	ATGAGTACTC	TGGGGCTTTA	TTTGGTCATC	ATCTTTTGTG	CCGCACAGTT
	1101	TGTCGCATTT	TTTAATTGGA	CGAATATTGG	GCAATATATT	GCCGTAAAG
	1151	GGGCGACGTT	CTTAAAGAA	GTCGGCTTGG	GCGGCAGCGT	GTTGTTTATC
	1201	GGTTTTATTT	TAATTTGTGC	TTTATCAAT	CTGATGATAG	GCTCCGCCTC
35	1251	CGCGCAATGG	GCGGTAAGTC	CGCCGATTTT	CGTCCCTATG	CTGATGTTGG
	1301	CCGGCTACGC	GCCCGAAGTC	ATTCAAGCCG	CTTACCGCAT	CGGTGATTCC
	1351	GTTACCAATA	TTATTACGCC	GATGATGAGT	TATTTCCGGC	TGATTATGGC
	1401	GACGGTGATC	AAATACAAAA	AAGATGCGGG	CGTGGGTACG	CTGATTCTTA
	1451	TGATGTTGCC	GTATTCCGCT	TTCTTCTTGA	TTGCGTGGAT	TGCCTTATTC
40	1501	TGCATTTGGG	TATTTGTTTT	GGGCCTGCC	GTCGGTCCCG	GCGCGCCAC
	1551	ATTCTATCCC	GCACCTTAA			

This encodes a protein having amino acid sequence <SEQ ID 138>:

	1	MSQTDTRQDG	RFLRTVEWLG	NMLPHPVTLF	IIFIVLLLIA	SAAGAYFGLS
	51	VPDRPVGAK	GRADDGLIHV	VSLLDADGLI	KILTHTVKNF	TGFAPLGTVL
45	101	VSLLGVGIAE	KSGLISALMR	LLLTSPKRL	TTFMVFTGI	LSNTASELGY
	151	VVLIPLSAII	FHSLGRHPLA	GLAAAFAGVS	GGYSANFLG	TIDPLLAGIT
	201	QQAQTIHPD	YVVGPEANWF	FMVASTFVIA	LIGYFVTEKI	VEPQLGPYQS
	251	DLSQEEKDIR	HSNEITPLEY	KGLIWAGVVF	VALSALLAWS	IVPADGILRH
	301	PETGLVSGSP	FLKSIVVFIF	LLFALPGIVY	GRVTRSLRGE	QEVVNAMAES
50	351	MSTLGLYLVI	IFFAAQFVAF	FNWTNIGQYI	AVKGATFLKE	VGLGGSVLF
	401	GFILICAFIN	LMIGSASAQW	AVTAPIFVPM	LMLAGYAPEV	IQAAAYRIGDS
	451	VTNIITPMMS	YFGLIMATVI	KYKKGAGVGT	LISMMLPYSA	FFLIAWIALF
	501	CIWVFLGLP	VPGGAPTFYP	AP*		

55 ORF12a and ORF12-1 show 99.0% identity in 522 aa overlap:

		10	20	30	40	50	60
	orf12a.pep	MSQTDTRQDG	RFLRTVEWLG	NMLPHPVTLF	IIFIVLLLIA	SAAGAYFGLS	VPDRPVGAK
60	orf12-1	MSQTDTRQDG	RFLRTVEWLG	NMLPHPVTLF	IIFIVLLLIA	SAVAGAYFGLS	VPDRPVGAK
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf12a.pep	GRADDGLIHV	VSLLDADGLI	KILTHTVKNF	TGFAPLGTVL	VSLLGVGIAE	KSGLISALMR
65	orf12-1	GRADDGLIYV	SLLDADGFI	KILTHTVKNF	TGFAPLGTVL	VSLLGVGIAE	KSGLISALMR

-131-

		70	80	90	100	110	120
5	orf12a.pep	130	140	150	160	170	180
	orf12-1	130	140	150	160	170	180
10	orf12a.pep	190	200	210	220	230	240
	orf12-1	190	200	210	220	230	240
15	orf12a.pep	250	260	270	280	290	300
	orf12-1	250	260	270	280	290	300
20	orf12a.pep	310	320	330	340	350	360
	orf12-1	310	320	330	340	350	360
25	orf12a.pep	370	380	390	400	410	420
	orf12-1	370	380	390	400	410	420
30	orf12a.pep	430	440	450	460	470	480
	orf12-1	430	440	450	460	470	480
35	orf12a.pep	490	500	510	520		
	orf12-1	490	500	510	520		

45 Homology with a predicted ORF from *N.gonorrhoeae*ORF12 shows 92.5% identity over a 320aa overlap with a predicted ORF (ORF12.ng) from *N.**gonorrhoeae*:

50	orf12.pep	AXXIIHPXXVVGPEANWFFMVASTFVIALI	30
	orf12ng	AAAFAGVSGGYSANLFLGTIDPLLAGITQAAQIIHPDYVVGPEANWFFMAASTFVIALI	232
55	orf12.pep	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV	90
	orf12ng	GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV	292
60	orf12.pep	PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS	150
	orf12ng	PADGILRHPETGLVAGSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNMAESMS	352
65	orf12.pep	TLXLXLXIIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIFIGFILICAFINLM	210
	orf12ng	TLGLYLVIIFFAAQFVAFFNWTNIGQYIAVKGAVFLKKFRLGGSVLFIFIGFILICAFINLM	412
	orf12.pep	IGSASQWAVTAPIFVPMMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVXXY	270
	orf12ng	IGSASQWAVTAPIFVPMMLAGNAPQVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKY	472

-132-

```

orf12.pep      KKDAGVGTLIXMMLPYSAFFLIWIALFCIWVFLGLPVGPGAPTFFYPAP 320
               |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12ng        KKDAGVGTLISMLPYSAFFLIWIALFCIWVFLGLPVGPGTPTFFYPVP 522

```

The complete length ORF12ng nucleotide sequence <SEQ ID 139> is:

```

5      1  ATGAGTCAAA CCGACGCGCG TCGTAGCGGA CGATTTTAC GCACAGTCGA
51     51  ATGGCTGGGC AATATGTTGC CGCACC CGGT TACGCTTTT ATTATTTTCA
101    101  TTGTGTTATT GCTGATTGcc tctgCCGTCG GTGCGTATT CGGACTATCC
151    151  GTCCCGGATC CGCGTCCTGT TGGGGCGAAA GGACGTGCCG ATGACGGTTT
201    201  GATTCACGTT GTCAGCCTGC TCGATGCCGA CGGTTTGATC AAAATCCTGA
10     251  CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCGGTTGGG AACGGTGTG
301    301  GTTTCCTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTCCGCG
351    351  ATTAATGCGC TTATTGCTCA CAAATCCCC ACGCAAATC ACTACTTTTA
401    401  TGGTTGTTTT TACAGGGATT TTATCCAATA CGGCTTCTGA ATTGGGCTAT
451    451  GTCGTCTCTAA TCCCTTTGTC CGCGTCATC TTTCATTGCG TCGGCCGCCA
15     501  TCCGCTTGCC GGTGTTGGCTG CGGCTTTCGC CGGCGTTTCG GCGGGTTAT
551    551  CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTTGGC AGGCATCACC
601    601  CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
651    651  CAACTGGTTT TTTATGGCAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
701    701  ATTTTGTTAC TGAAAAATC GTCGAACC GC AATTGGGCCC TTATCAATCA
20     751  GATTGTGCAC AAGAAGAAAA AGACATTCGG CATTCCAATG AAATCACGCC
801    801  TTTGGAATAT AAAGGATTAA TTTGGGCAGG CGTGGTGT TTGCTTAT
851    851  CCGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
901    901  CCTGAAACAG GATTGGTTGC CGGTTCGCCG TTTTAAAAAT CGATTGTTGT
951    951  TTTTATTTTC TTGTTGTTTG CGCTGCCGGG CATTGTTTAT GGCCGGATAA
25     1001  CCCGAAGTTT GCGCGGCGAA CGGGAAGTCG TTAATGCGAT GGCCGAATCG
1051   1051  ATGAGTACTT TGGGACTTTA TTTGGTCATC ATCTTTTTTG CCGCACAGTT
1101   1101  TGTCGCATT TTTAATTGGA CGAATATTGG GCAATATATT GCCGTAAAG
1151   1151  GGGCGGTGTT CTTAAAGAA GTCGGCTTGG GCGGCAGTGT GTTGTATC
1201   1201  GGTTTTATTT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCTC
30     1251  CGCGCAATGG GCGGTAATG CGCGGATTTT CGTCCCTATG CTGATGTTGG
1301   1301  CCGGCTACGC GCCCGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
1351   1351  GTTACCAATA TTATTACGCC GATGATGAGT TATTTCGGGC TGATTATGGC
1401   1401  GACGGTAATC AAATACAAAA AAGATGCGGG CGTAGGCACG CTGATTCTA
1451   1451  TGATGTTGCC GTATTCGCT TTCTTCTTAA TTGCATGGAT CGCCTTATC
35     1501  TGCATTTGGG TATTTGTTT GGGTCTGCCC GTCGGTCCCG GCACACCCAC
1551   1551  ATTCTATCCG GTGCCTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 140>:

```

1  MSQTDARRSG RFLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAVGAYFGLS
51  VPDPRPVGAK GRADDGLIHV VSLLDADGLI KILHTVKNF TGFAPLGTVL
40  101  VSLLGVGIAE KSLGISALMR LLLTKSPRKL TTFMVVFTGI LSNTASELGY
151  VVLIPLSAVI FHSLSGRHPLA GLAAAFAGVS GGYSANFLG TIDPLLAGIT
201  QAAQIIHPD YVVGPEANWF FMAASTFVIA LIGYFVTEKI VEPQLGPYQS
251  DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
301  PETGLVAGSP FLKSIVVFIF LLFALPGIVY GRITRSLRGE REVVNMAES
45  351  MSTLGLYLVI IFFAAQFVAF FNWTNIGQYI AVKGAVFLKK FRLGGSVLF
401  GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGNAPQV IQAAYRIGDS
451  VTNIITPMMS YFGLIMATVI KYKKDAGVGT LISMLPYSA FFLIWIWIALF
501  CIWVFLGLP VPGTPTFFYP VP*

```

ORF12ng shows 97.1% identity in 522 aa overlap with ORF12-1:

```

50      10      20      30      40      50      60
orf12-1.pep  MSQTDQDGRFLRTVEWLG NMLPHPVTLF IIFIVLLLIASAVGAYFGLSVDPDRPVGAK
               |||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12ng      MSQTDARRSGRFLRTVEWLG NMLPHPVTLF IIFIVLLLIASAVGAYFGLSVDPDRPVGAK
               10      20      30      40      50      60

55      70      80      90     100     110     120
orf12-1.pep  GRADDGLIYIVSLLNADGFIKILHTVKNF TGFAPLGTVLVSLLGVGIAEKSLGISALMR
               |||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12ng      GRADDGLIHVVSLLDADGLIKILHTVKNF TGFAPLGTVLVSLLGVGIAEKSLGISALMR
               70      80      90     100     110     120

60      130     140     150     160     170     180
orf12-1.pep  LLLTKSPRKL TTFMVVFTGILSNTASELGYVVLIPLSAIIFHSLSGRHPLAGLAAAFAGVS
               |||||:|||||:|||||:|||||:|||||:|||||:|||||:
orf12ng      LLLTKSPRKL TTFMVVFTGILSNTASELGYVVLIPLSAIIFHSLSGRHPLAGLAAAFAGVS
               130     140     150     160     170     180

```

-133-

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf12-1.pep	GGYSANLFLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMVASTFVIALIGYFVTEKI					
	orf12ng	GGYSANLFLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKI					
10	orf12-1.pep	250	260	270	280	290	300
	orf12ng	250	260	270	280	290	300
15	orf12-1.pep	310	320	330	340	350	360
	orf12ng	310	320	330	340	350	360
20	orf12-1.pep	370	380	390	400	410	420
	orf12ng	370	380	390	400	410	420
25	orf12-1.pep	430	440	450	460	470	480
	orf12ng	430	440	450	460	470	480
30	orf12-1.pep	490	500	510	520		
	orf12ng	490	500	510	520		

In addition, ORF12ng shows significant homology with a hypothetical protein from *E.coli*:

40	sp P46133 YDAH_ECOLI HYPOTHETICAL 55.1 KD PROTEIN IN OGT-DBPA INTERGENIC REGION	
	>gi 1787597 (AE000231) hypothetical protein in ogt 5'region [Escherichia coli]	
	Length = 510	
	Score = 329 bits (835), Expect = 2e-89	
	Identities = 178/507 (35%), Positives = 281/507 (55%), Gaps = 15/507 (2%)	
45	Query: 8	RSGRFLRTVEWLG NMLPHPV TXXXXXXXXXXASAVGAYFGLSVPDPRPVGAKGRADDGL 67
		+SG+ VE +GN +PHP +A+ +FG+S +P D
	Sbjct: 13	QSGKLYGWVERIGNKVPHPFLFIYLIIVLMVTTAILSAFGVSAKNP-----TDGTP 64
50	Query: 68	IHVVSLLDADGLIKILTHTVKNFTGFAPXXXXXXXXXXIAEKSLISALMRLLLT KSP 127
		+ V +LL +GL L + +KNF+GFAP +AE+ GL+ ALM + +
	Sbjct: 65	VVVKNLLSVEGLHWFLPNVIKNFSGFAPLGAILALVLGAGLAERVGLLPALMVKMASHVN 124
55	Query: 128	RKLTTFMVVF TGILSNTASELG YVVLIPLSAVIFHSLGRHPLAGLAAAFAGVSGGYSANL 187
		+ ++MV+F S+ +S+ V++ P+ A+IF ++GRHP+AGL AA AGV G++ANL
	Sbjct: 125	ARYASYMVLFI AFFSHISSDAALVIMPPMGALIFLAVGRHPVAGLLAAIAGVCGCFTANL 184
60	Query: 188	FLGTIDPLLAGITQQAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKIVEPQLGP 247
		+ T D LL+GI+ +AA +P V NW+FMA+S V+ ++G +T+KI+EP+LG
	Sbjct: 185	LIVTTDVLLSGISTEAAA AFNPQMHVSVIDNWYFMASVVVLTIVGGLITDKIIEPRLGQ 244
65	Query: 248	YQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRHPETGLVA 307
		+Q + ++ + + S GL AGVV + A +A ++P +GILR P V
	Sbjct: 245	WQGNSDEKLQTLTESQRF-----GLRIAGVVSLLFIAAIALMVIPQNGILRDPINHTVM 298
70	Query: 308	GSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAMAESMSTLGLYLXXXXXXX 367
		SPF+K IV I L F + + YG TR++R + ++ + M E M + ++
	Sbjct: 299	PSPFIKGIVPLIILFFVVS LAYGIATRTIRRQADLPHLMIEPMKEMAGFIVMVFP LAQF 358
	Query: 368	XXXXNWTNIGQYIAVKGAVFLKEVGLGGSVLFIFIGFILICAFINLMIGSASAQWAVTAPIF 427
		NW+N+G++IAV L+ GL G F+G L+ +F+ + I S SA W++ APIF

-134-

Sbjct: 359 VAMFNWSNMGKFIAGVGLTDILESSGLSGIPAFVGLALLSSFLCMFIASGSASISILAPIF 418
 Query: 428 VPMLMLAGYAPEVIQAAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGTLSMMLP 487
 VPM ML G+ P Q +RI DS + P+ + L + + +YK DA +GT S++LP
 5 Sbjct: 419 VPMFMLLGfHPAFAQILFRIADSSVLPLAPVSPFVPLFLGLQRYKPDALGTYISLVLP 478
 Query: 488 YSAFFLIAWIALFCIWFVVLGLPVGPG 514
 Y FL+ W+ + W +++GLP+GPG
 10 Sbjct: 479 YPLIFLVVWLLMLLAW-YLVGLPIGPG 504

Based on this analysis, including the presence of several putative transmembrane domains and the predicted actinin-type actin-binding domain signature (shown in bold) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

15 Example 17

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 141>:

1 ..ACAGCCGGCG CAGCAGGTTn CnCGGTCTTC GTTTTCGTAA CGGACAGTCA
 51 GGTGGAGGTG TTCGGGAACA TCCAGACCGC AGTGGAACA GGTTTTTTTC
 101 ATGGCATTTC GGTTCGTCT GTGTTTGGTG CGGCGGCACA AGACTCGGCA
 151 ATgGCTTCGC GCAGTGCCTC TATACCGGTA TTTTCAGCAA CGGAAATGCG
 201 GACGGcGgCA ATTTTCCCG CAGCGTCGCG CCATATGCCG GTGTTTgTT
 251 CTTCAGACGG CAGCAGGTCG GTTTTGTGT ACACCTTgAT GCACGGAaTA
 301 TCGCCGGCAT GGATTCTTG CAGTACGTT TCCACGTCTT CAATCTGCTG
 351 TCCGCTGTTC GGAGCGGCGG CATCGACGAC GTGCAGCAGC ACATCgGcTT
 25 401 gCGCGGTTTC TTCCAGCGTG GCgGAAAAGG CGGAAATCAG TTTgTGCggc
 451 agATyGCTnA CGAATCCGAC GGTATCGGTC AGGATAATGC TGCATTCGGG
 501 ACT..

This corresponds to the amino acid sequence <SEQ ID 142; ORF14>:

1 ..TAGAAGXXVF VFTDSQVEV FGNIQTAVET GFFHGISVSS VFGAAQDSA
 30 51 MASRSASIPV FSATEMRTAA IFPAASRHMP VFCSSDGSRs VLLYTLMHGI
 101 SPAWISCSTF STSSICPLF GAAASTCCS TSACAVSSSV AEKAEISLCG
 151 RXLTNPTVSV RIMLHSG..

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

35 ORF14 shows 94.0% identity over a 167aa overlap with an ORF (ORF14a) from strain A of *N.meningitidis*:

	orf14.pep		10	20	30
			TAGAAGXXVFVFTDSQVEVFGNIQTAVET		
40	orf14a	GRQLGFLRVGGALFVITAQARVNALCDCLTTGAAGFAVFVFTDGQMQVFGNVQPAVET	150	160	170
			180	190	200
	orf14.pep		40	50	60
45			70	80	90
	orf14a	GFFHGISVSSVFGAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRs	210	220	230
			240	250	260
50	orf14.pep	VLLYTLMHGISPAWISCSTFSTSSICPLFGAAASTCCSSTSACAVSSSVAEKAEISLCG	100	110	120
	orf14a	VLLYTLMHGISPAWISCSTFSTSSICPLFGAAASTCCSSTSACAVSSSVAEKAEISLCG	130	140	150
			270	280	290
			300	310	320

```

                                160
orfl4.pep    RXLTNPVTSVRIMLHSG
              | | | | | | | | | |
orfl4a       RSLTNPTVTSVRIMLHSGLMYSRRRAVVSSVAKSWSFAYMPDLVSRLNRLDLPTLVX
5            330      340      350      360      370      380

```

The complete length ORF14a nucleotide sequence <SEQ ID 143> is:

	1	ATGGAGGATT	TGCAGGAAAT	CGGGTTCGAT	GTCGCCGCCG	TAAAGGTAG
	51	TCGCGACGCG	GAACATCATC	GTCTGCATCA	TCCCCAGCCC	GGCAACGGCG
	101	AGGCGGACGA	TGTATTGTTT	GCGTTCTTTT	TGGTTGGCGG	CTTCGATTTT
10	151	TTGCCGCGTA	TAGGGTGGCG	CGGTGTAGCC	TATCTGCCTT	ATTTTCAACA
	201	GAATGTCGGA	AAGCGGATT	TTGCCGTCGT	CCCGACGACG	CGCGGACGGG
	251	TCGGTGCTGT	AATTGAGGTC	GATGCGGACG	ATGCCGTCGT	TACGCAAAAG
	301	CTGCTGTTTC	ATCAGCCAGA	CGCAGGCGGC	GCAGGTGATG	CCGCGCGAGCA
	351	TTAAACC	CTCGCGCGTG	CCGCGGTGGG	TTTCCACAAA	GTCGGAGCTGG
15	401	ACTTCCGGCA	GGTCGTACAG	GCGGATTTGG	TCGAGGATTT	CTTGGGGCGG
	451	CAGCTCGGTT	TTTTGCGCGT	CGGCGGTGCG	TTGTTTGTAA	TAACTGCCCA
	501	AGCCCGCGTC	AATAATGCTT	TGTGCGACTG	CCTGACAACC	GGCGCAGCAG
	551	GTTTTCGCGGT	CTTCGTTTTT	GTAACGGACG	GTCAGATGCA	GGTTTTTCGG
	601	AACGTCACG	CCGCACTGGA	AACAGGTTTT	TTTCATGGCA	TTTCGGTTTT
20	651	GCTGTGTTTT	GGTCGCGCGG	CACAATACTC	GGCAATGGCT	TCGCGCAGTG
	701	CGTCTATACC	GGTATTTTCA	GCAACGGAAA	TGCGGACGGC	GGCAATTTTT
	751	CCCGCAGCGT	CGCGCCATAT	GCCCCGTGTT	TGTTCTTCAG	ACGGCAGCAG
	801	GTCGGTTTTG	TTGTACACCT	TGATGCACGG	AATATGCCCG	GCATGGATTT
	851	CTTGACGTAC	TTTTTCCACG	TCTTCAATCT	GCTGTCCGCT	GTTCCGGAGCG
25	901	GCGGCATCGA	CGACGTGCAG	CAGCACATCG	GCTTGCGCGG	TTTCTTCCAG
	951	CGTGGCGGAA	AAGGCGGAAA	TCAGTTTGTG	CGGCAGATCG	CTGACGAATC
	1001	CGACGGTATC	GGTCAGGATA	ATGCTGCATT	CGGGACTGAT	GTACAGCCCG
	1051	CGCGCCGTCG	TGTCGAGTGT	GCGCAAAAGC	TGGTCTTTTC	CATATATGCC
	1101	CGAATTGGTC	AGCCGGTTGA	ACAGACTGGA	TTTGCCGACA	TTGGTATAG

30 This encodes a protein having amino acid sequence <SEQ ID 144>:

	1	MEDLQEIGFD	VAAVKVGRQR	EHHRLHHPQP	GNGEADDVLF	AFFLVGGFDF
	51	LRVIGCGGVA	YLPDFQQNVG	KADFAVVPDD	AAAVRAVIEV	DADDAVCTQK
	101	LLFDQPDAGG	AGDAAEH*NR	LARAAVGFHK	VGLDFGQVVQ	ADLVEFDLGR
	151	QLGLFLRVGGA	LFVITAQARV	NNALCDCLTT	GAGAFVVFVF	VTDGQMQVFG
35	201	NVQPAVETGF	FHGISVSSVF	GAAQYSAMA	SRSASIPVFS	ATEMRTAAIF
	251	PAASRHMPVF	CSSDGSRSVL	LYTLMHGISP	ATWISCTFST	SSICCPLEGA
	301	AASTTCSSTS	ACSAVSSVAE	KAEISLCGRS	LTNPVTSVRI	MLHSGLMYSR
	351	RAVVSSVAKS	WCFAYMPDLV	SRLNRLDLPT	LV*	

It should be noted that this sequence includes a stop codon at position 118.

40 Homology with a predicted ORF from *N.gonorrhoeae*

ORF14 shows 89.8% identity over a 167aa overlap with a predicted ORF (ORF14.ng) from *N.*

gonorrhoeae:

	orf14.pep	TAGAAGXXVVFVTDSEQVEVFGNIQTAVET	30
		: : : : :	
45	orf14.ng	GRQFGFFRVGGASFVITAQAGIDDALCCLTADAAGFAVFAFVADGQMQVEFGNVQPAVET	208
	orf14.pep	GGFHGISVSSVFGAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHPVFCSSDGSRS	90
50	orf14.ng	GGFHGISVSSVFGAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHPVFCSSDGSRS	268
	orf14.pep	VLLYTLMHGISPAWISCSTFSTSSICPLFGAAASTTCSSTSACAVSSSVAEKAIEISLCG	150
	orf14.ng	VLLYTLMHGISWAWISCSTFSTSSICPLFRAAASTTCSSTSACTVSSKVAEKAIEISLCG	328
55	orf14.pep	RXLTNPTVSVRIMLHSG	167
		:	
	orf14.ng	RSLTNPTVSVRIMLHAGLMYSRRVAVSVRAKSWSFAYMPDLVSRNLNRLDLPTLV	382

The complete length ORF14ng nucleotide sequence <SEQ ID 145> is predicted to encode a protein having amino acid sequence <SEQ ID 146>:

1 MEDLQEIGFD VAAVKVGRQR EHHRLHHTQS GNGKADDVLF AFFLVGGFDF
 51 LRVIGCGGVA CLPDFQONVG EADFAVVPDD AAAVRAVIEV DADDAVCAQK
 101 LLFDQPDAGG AGNAAEHQHC FVRAIMGFHK VGLDFGQVVO ADLVEDFLGR
 151 QFGFFRVGGA SFVITAQAGI DDALCDCLTA DAAGFAVFAF VADGQMQVFG
 201 NVQPAVETGF FHGISVSSVF GAAQYSAMA SRSASIPVFS ATEMRTAAIF
 251 PAASRHPVF C SSDGSRSVL LYTLMHGISW AWISCSTFST SSICCPLFRA
 301 AASTTCSSTS ACTVSSKVAE KAEISLCGRS LTNPTVSVRI MLHAGLMYSR
 351 RAVVSRVAKS WSFAYMPDLV SRLNRLDLPT LV*

Based on the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 18

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 147>:

1 ..GGCCATTACT CCGACCGCAC TTGGAAGCCG CGTTTGGNCG GCCGCCGTCT
 51 GCCGTATCTG CTTTATGGCA CGCTGATTGC GGTATTGTG ATGATTTTGA
 101 TGCCGAACCTC GGGCAGCTTC GGTTCGGCT ATGCGTCGCT GCGCGCTTTG
 151 TCGTTCGGCG CGCTGATGAT TCGCTGTGA GACGTGTCGT CAAATATGGC
 201 GATGCAGCCG TTAAAGATGA TGGTCGGCGA CATGGTCAAC GAGGAGCAGA
 251 AAA.NTACGC CTACGGGATT CAAAGTTTCT TAGCAAATAC GGGCGCGGTC
 301 GTGGCGGCGA TTCTGCCGTT TGTGTTTGGC TATATCGGTT TGGCGAACAC
 351 CGCCGANAAA GCGTGTGTGC CGCAGACCGT GGTCTGGCG TTTTATGTGG
 401 GTGCGGCGTT GCTGGTGATT ACCAGCGCGT TCACGATTTT CAAAGTGAAG
 451 GAATACGANC CGGAAACCTA CGCCCGTTAC CACGGCATCG ATGTCGCCGC
 501 GAATCAGGAA AAAGCCAACT GGATCGCACT CTAAAA.CC GCGC..

25 This corresponds to the amino acid sequence <SEQ ID 148; ORF16>:

1 ..GHYSDRTWKP RLXGRRLPYL LYGTLIAVIV MILMPNSGSF GFGYASLAAL
 51 SFGALMIALI DVSSNMAMQP FKMMVGDVMN EEQKXYAYGI QSFLANTGAV
 101 VAAILPFVFA YIGLANTAXK GVPQTVVVA FVVGAAALVI TSAFTIFVKV
 151 EYXPETYARY HGIDVAANQE KANWIALLLKX A..

30 Further work revealed the complete nucleotide sequence <SEQ ID 149>:

1 ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC
 51 AAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG
 101 CCTTTACCTT GCAAAGCTCG CAAATGAGCC GCATTTTCA AACGCTAGGC
 151 GCAGACCCGC ACAATTGGG CTGGTTTTTC ATCCTGCCGC CGCTGGCGGG
 201 GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGAAGC
 251 CGCGTTTGGG CGGCCGCGT CTGCCGTATC TGCTTTATGG CACGCTGATT
 301 GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG
 351 CTATGCGTCG CTGGCGGCTT TGTCGTTCCG CGCGCTGATG ATTGCGCTGT
 401 TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC
 451 GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
 501 CTTAGCAAAT ACGGGCGCGG TCGTGGCGGC GATTCTGCCG TTTGTGTTTG
 551 CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGCAGACC
 601 GTGGTCGTGG CGTTTTATGT GGGTGCGGCG TTGCTGGTGA TTACCAGCGC
 651 GTTACGATT TTCAAAGTGA AGGAATACGA TCCGGAACC TACGCCCGTT
 701 ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA
 751 CTCTTGAAAA CCGCGCTTAA GGCGTTTGG ACGGTTACTT TGGTGCAATT
 801 CTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA
 851 TTGCGGAAAA CGTCTGGCAC ACCACGATG CGTCTCCGT AGGTTATCAG
 901 GAGGCGGGTA ACTGGTACGG CGTTTGGCG GCGGTGCACT CGGTTGCGCG
 951 GGTGATTTGT TCGTTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG
 1001 CGGGTTATTT CGGCTGTTTG GCTTGGGCG CGCTCGGCTT TTTCTCCGTT
 1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGCTTATA CCTTAATCGG
 1101 CATCGCTTGG GCGGGCATTA TCACTTATCC GCTGACGATT GTGACCAACG
 1151 CCTGTGCGGG CAAGCATATG GGCACCTACT TGGGCTTGT TAACGGCTCT
 1201 ATCTGTATGC CTCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC
 1251 TATGCTGGGC GGCTTGCAGG CCACTATGTT CTTGGTAGGG GCGCTCGTCC
 1301 TGCTGCTGGG CGCGTTTCC GTGTTCTGTA TTAAGAAAC ACACGGCGGG
 1351 GTTTGA

This corresponds to the amino acid sequence <SEQ ID 150; ORF16-1>:

```

      1 MSEYTPQAK OGLPALAKST IWMLSFGLG VQTAFTLQSS QMSRIFQTLG
    51 ADPHNLGWFF ILPPLAGMLV QPIVGHYSR TWKPRLGGR LPYLLYGTLI
   101 AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG
   151 DMVNEEQGY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVVPQT
   201 VVVAFYVGAA LLVITSFTI FKVKEYDPET YARYHGIDVA ANQEKANWIE
   251 LLKTAPKAFW TVTLVQFFCW FAFQYMWYTS AGAIAENVWH TTDASSVGYY
   301 EAGNWWGVLA AVQSVAAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV
   351 FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGFLNGS
   401 ICMPOIVASL LSFVLFPMGLG GLQATMFLVG GVVLLLAGFS VFLIKETHGG
   451 V*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF16 shows 96.7% identity over a 181aa overlap with an ORF (ORF16a) from strain A of *N.*

15 *meningitidis*:

```

                                     10      20      30
    orf16.pep                      GHYSDRTWKPRLXGRRLPYLLYGTLIIV
                                     |||
   20 orf16a                      IFQTLGADPHSLGWFFILPPLAGMLVQPIVGHYSR
                                     50      60      70      80      90     100
                                     |||
                                     40      50      60      70      80      90
   25 orf16a                      MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGMVNEEQKXYAYGI
                                     110     120     130     140     150     160
                                     |||
                                     100     110     120     130     140     150
   30 orf16a                      QSFLANTGAVVAAILPFVAYIGLANTAXKGVVPQTVVVAFYVGAAALLVITSFTIFKVK
                                     170     180     190     200     210     220
                                     |||
                                     160     170     180
   35 orf16a                      EYXPETYARYHGIDVAANQEKANWIALKXA
                                     ||
   40 orf16a                      EYNPETYARYHGIDVAANQEKANWIELKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAI
                                     230     240     250     260     270     280
                                     |||
   40 orf16a                      AENVVHTTDASSVGYYEAGNWWGVLAQSVAAVICSFVLAKVPNKYHKAGYFGCLALGA
                                     290     300     310     320     330     340

```

The complete length ORF16a nucleotide sequence <SEQ ID 151> is:

```

      1 ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC
    51 AAAAAGCAGC ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG
   101 CCTTTACCCT GCAAAGCTCG CAGATGAGCC GCATCTTCCA GACGCTCGGT
   151 GCCGATCCGC ACAGCCTCGG CTGGTTCTTT ATCCTGCCGC CGCTGGCGGG
   201 GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGAAGC
   251 CGCGTTGGG CGGCCGCCGT CTGCCGTATC TGCTTTATGG CACGCTGATT
   301 GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG
   351 CTATGCGTCG CTGGCGGCTT TGTCGTTCGG CGCGCTGATG ATTGCGCTGT
   401 TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC
   451 GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
   501 CTTAGCGAAT ACGGGCGCGG TCGTGGCGGC GATTCTGCCG TTTGTGTTTG
   551 CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCCGAGACC
   601 GTGGTCGTGG CGTTTATGT GGGTGGCGCG TTGCTGGTGA TTACCAGCGC
   651 GTTCACGATT TTCAAAGTGA AGGAATACAA TCCGGAAACC TACGCCGTT
   701 ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA
   751 CTCTTGAAAA CCGCGCCTAA GCGGTTTGG ACGGTTACTT TGGTGCAATT
   801 CTTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA
   851 TTGCGGAAAA CGTCTGCAC ACCACCGATG CGTCTCCGT AGGTATCAG
   901 GAGGCGGGTA ACTGGTACGG CGTTTGGCG GCGGTGAGT CGGTTGCGGC
   951 GGTGATTGT TCGTTGTAT TGGCGAAGT GCCGAATAAA TACCATAAGG

```


-138-

5
 1001 CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT
 1051 TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG
 1101 CATCGCTTGG GCGGGCATTG TCACTTATCC GCTGACGATT GTGACCAACG
 1151 CCTTGTGCGG CAAGCATATG GGCACCTACT TGGGCCTGTT TAACGGCTCT
 1201 ATCTGTATGC CGCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC
 1251 TATGCTGGGC GGCTTGCAGG CCACTATGTT CTTGGTAGGG GGCCTCGTCC
 1301 TGCTGCTGGG CGCGTTTTCC GTGTTCTCTGA TTAAAGAAAC ACACGGCGGG
 1351 GTTTGA

This encodes a protein having amino acid sequence <SEQ ID 152>:

10
 1 MSEYTPQAK QGLPALAKST IWMLSFGFLG VQTAFTLQSS QMSRIFQTLG
 51 ADPHSLGWFF ILPPLAGMLV QPIVGHYSR TWKPRLGRR LPYLLYGTLI
 101 AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG
 151 DMVNEEQGY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVVPQT
 201 VVAFYVGAA LLVITSFTI FKVKEYNPET YARYHGIDVA ANQEKANWIE
 15
 251 LLKTAPKAFW TVTLVQFFCW FAFQYMWYTS AGAIAENVWH TTDASSVGQY
 301 EAGNWWYGVLA AVQSVAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFESV
 351 FFIGNOYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGFLNGS
 401 ICMPQIVASL LSFVLFPMLG GLQATMFLVG GVVLLGAFS VFLIKETHGG
 451 V*

20 ORF16a and ORF16-1 show 99.6% identity in 451 aa overlap:

		10	20	30	40	50	60
orfl6a.pep		MSEYTPQAKQGLPALAKSTIWMLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHSLGWFF					
orfl6-1		MSEYTPQAKQGLPALAKSTIWMLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFF					
		10	20	30	40	50	60
orfl6a.pep		70	80	90	100	110	120
orfl6-1		70	80	90	100	110	120
		130	140	150	160	170	180
orfl6a.pep		LAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQGYAYGIQSFLANTGAVVAAILP					
orfl6-1		LAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQGYAYGIQSFLANTGAVVAAILP					
		130	140	150	160	170	180
orfl6a.pep		190	200	210	220	230	240
orfl6-1		190	200	210	220	230	240
		250	260	270	280	290	300
orfl6a.pep		ANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAIAENVWHHTDASSVGQY					
orfl6-1		ANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWYTSAGAIAENVWHHTDASSVGQY					
		250	260	270	280	290	300
orfl6a.pep		310	320	330	340	350	360
orfl6-1		310	320	330	340	350	360
		370	380	390	400	410	420
orfl6a.pep		LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTLYLGLFNGSICMPQIVASLLSFVLFPMLG					
orfl6-1		LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTLYLGLFNGSICMPQIVASLLSFVLFPMLG					
		370	380	390	400	410	420
orfl6a.pep		430	440	450			
orfl6-1		430	440	450			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF16 shows 93.9% identity over a 181aa overlap with a predicted ORF (ORF16.ng) from *N.*

gonorrhoeae:

5	orf16.pep	GHYSDRTWKPRXLGRRLLPYLLYGTLIAVIV	30
	orf16ng	HFNSNARRRPAQFGLVFHPAAAGGDAGSADSGYSDRTWKPRLGRRLLPYLLYGTLIAVIV	131
10	orf16.pep	MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQKXYAYGI	90
	orf16ng	MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDVNEEQKXYAYGI	191
	orf16.pep	QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQTVVVAFYVGAALLVITSFTISKVK	150
15	orf16ng	QSFLANTDAVVAAILPFVFAYIGLANTAEGVVPQTVVVAFYVGAALLIITSFTISKVK	251
	orf16.pep	EYXPETYARYHGIDVAANQEKANWIALLKXA	181
	orf16ng	EYDPETYARYHGIDVAANQEKANWFELLKTAPKVFVTVPVQFFCWFAPRYMWTYSAGAI	311

20 The complete length ORF16ng nucleotide sequence <SEQ ID 153> is:

1	ATGATAGGGG	ATCGCCGCGC	CGGCAACCAT	TTCGGATTTT	CCAAAGCAAA
51	TACTTTTCAA	ATCAAAAAAA	AGGATTTACT	TTATGTCGGA	ATATACGCCT
101	CAACAGCAA	AACAAGGTTT	GCCCGCGCCG	GCAAAAAGCA	CGATTGGAT
151	GTTGAGCTTC	GGCTATCTCG	GCGTTCAGAC	GGCCTTTACC	CTGCAAAGCT
201	CGCAGATGAG	CCGCATTTT	CAACGCTAG	GCGCAGACCC	GCACAATTTG
251	GGCTGGTTT	TCATCCTGCC	GCCGCTGGCG	GGGATGCTGG	TTCAGCCGAT
301	AGTGCTACT	ACTCAGACCG	CACTTGGAG	CCGCGCTTGG	GCGGCGCGCG
351	CCTGCCGTAT	CTGCTTTACG	GCACGCTGAT	TGCGGTCATC	GTGATGATTT
401	TGATGCCGAA	CTCGGCGCAG	TTCGGTTTCG	GCTATGCGTC	GCTGGCGGCC
451	TTGTCGTTTC	GCGCGCTGAT	GATTCGCGTG	TTGGACGTGT	CGTCGAATAT
501	GGCGATGCAG	CCGTTTAAGA	TGATGGTCGG	CGATATGGTC	AACGAGGAGC
551	AGAAAAGCTA	CGCCTACGGG	ATTCAAAGTT	TCTTAGCGAA	TACGGACGCG
601	GTTGTGGCAG	CGATTCTGCC	GTTTGTGTTC	GCGTATATCG	GTTTGGCGAA
651	CACTGCCGAG	AAAGGCGTTG	TGCCACAAAC	CGTGGTCGTA	GCATTCTATG
701	TGGGTGCGGC	GTTACTGATT	ATTACCAGTG	CGTTCACAAT	CTCCAAAGTC
751	AAAGAATACG	ACCCGGAAC	CTACGCCCGT	TACCACGGCA	TCGATGTCGC
801	CGCGAATCAG	GAAAAGCCA	ACTGGTTCGA	ACTCTAAAA	ACCGCGCCTA
851	AAGTGTTTTG	GACGGTTACT	CCGGTACAGT	TTTTCTGCTG	GTTCCGCTTC
901	CGGTATATGT	GGACTTACTC	GGCAGGCGCG	ATTGCAGAAA	ACGTCTGGCA
951	CACTACCGAT	GCGTCTTCCG	TAGGCCATCA	GGAGGCGGGC	AACCGGTACG
1001	GCGTTTGGC	GCGGTGTAG			

This encodes a protein having amino acid sequence <SEQ ID 154>:

1	MIGDRRAGNH	FGFSKANTFQ	IKKKDLYVVG	IYASNSKTRF	ARAGKKHDL
51	VELRLSRSD	GLYPAKLADE	PHFSNARRRP	AQFGLVFHPA	AAGGDAGSAD
101	SGYSDRTWK	PRLGRRLLPY	LLYGTIAVI	VMILMPNSGS	FGFGYASLAA
151	LSFGALMIAL	LDVSSNMAMQ	PFKMMVGDV	NEEQKXYAYG	IQSFLANTDA
201	VVAAILPFVF	AYIGLANTAE	KGVPVQTVV	AFYVGAALLI	ITSFTISKV
251	KEYDPETYAR	YHGIDVAANQ	EKANWFELLK	TAPKVFVTVT	PVQFFCWFAP
301	RYMWTYSAGA	IAENVHHTD	ASSVGHQEAG	NRYGVLAAV*	

50 ORF16ng and ORF16-1 show 89.3% identity in 261 aa overlap:

		30	40	50	60	70	80
	orf16-1.pep	MLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFFILPPLAGMLVQPI-VGHYSDRT					
	orf16ng	DVELRLSRSDGLYPAKLADEPHFSNARRRPAQFGLVF-HPAAAGGDAGSADSGYSDRT					
55		50	60	70	80	90	100
	orf16-1.pep	WKPRLGRRLLPYLLYGTLIAIVMILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMA					
	orf16ng	WKPRLGRRLLPYLLYGTLIAIVMILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMA					
60		110	120	130	140	150	160

-140-

		150	160	170	180	190	200
	orf16-1.pep	MQPFKMMVGD	MVNNEEQKGYAYGI	QSF	LANTGAVVAAILP	PFVFAYIGL	ANTAEGVVPQTV
5	orf16ng	MQPFKMMVGD	MVNNEEQKSYAYGI	QSF	LANTDAVVAAILP	PFVFAYIGL	ANTAEGVVPQTV
		170	180	190	200	210	220
		210	220	230	240	250	260
	orf16-1.pep	VVAFYVGAALL	VITSFTIFKVKEYD	PETYARYHGID	VAAANQEKAN	WIELLKTAP	KAFWT
10	orf16ng	VVAFYVGAALL	IITSFTISKVKEYD	PETYARYHGID	VAAANQEKAN	WFELLKTAP	KVFWT
		230	240	250	260	270	280
		270	280	290	300	310	320
15	orf16-1.pep	VTLVQFFCWF	AFQYMWTSAGAIA	ENVWHTDASS	VGQYQ	EAGN	WYGVLA
	orf16ng	VTVPVQFFC	WFAFRYMWTSAGAIA	ENVWHTDASS	VGHQ	EAGN	RYGVLA
		290	300	310	320	330	340

- 20 Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 19

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 155>:

25	1	ATGTTGTTCC	GTAAACGAC	CGCCGCCGTT	TTGGCGCATA	CCTTGATGCT
	51	GAACGGCTGT	ACGTTGATGT	TGTGGGGAAT	GAACAACCCG	GTCAGCGAAA
	101	CAATCACCCG	NAAACACGTT	GNCAAAGACC	AAATCCGNGN	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAAAG	GGCAGCCTGG	TGATGATGGG
	201	CGGAAAATAC	TGGTTCGTCG	TCAATCCCGA	AGATTCGGCG	AA.NTGACGG
30	251	GNATTTTGAN	GGCAGGGCTG	GACAAACCCT	TCCAAATAGT	TNAGGATACC
	301	CCGAGCTATG	C.TGCCACCA	AGCCCTGCCG	GTCAAACCTG	GATCGNCTGG
	351	CAGCCAGAAT	...			

This corresponds to the amino acid sequence <SEQ ID 156; ORF28>:

35	1	MLFRKTTAAV	LAHTLMLNGC	TLMLWGMNPN	VSETITRKHV	XKDQIRXFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WFVVPEDSA	XTTGILXAGL	DKPFQIVXDT
	101	PSYXCHQALP	VKLGSXGSQN	...		

Further work revealed the complete nucleotide sequence <SEQ ID 157>:

	1	ATGTTGTTCC	GTAAACGAC	CGCCGCCGTT	TTGGCGGCAA	CCTTGATGCT
	51	GAACGGCTGT	ACGTTGATGT	TGTGGGGAAT	GAACAACCCG	GTCAGCGAAA
40	101	CAATCACCCG	CAAACACGTT	GACAAAGACC	AAATCCGCGC	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGAAAAG	GGCAGCCTGG	TGATGATGGG
	201	CGGAAAATAC	TGGTTCGTCG	TCAATCCCGA	AGATTCGGCG	AAGCTGACGG
	251	GCATTTTGAA	GGCAGGGCTG	GACAAACCCT	TCCAAATAGT	TGAGGATACC
	301	CCGAGCTATG	CTCGCCACCA	AGCCCTGCCG	GTCAAACCTG	AATCGCCTGG
45	351	CAGCCAGAAT	TTCAGTACCG	AAGGCCTTTG	CCTGCGCTAC	GATACCGACA
	401	AGCCTGCCCA	CATCGCCAAG	CTGAAACAGC	TCGGGTTTGA	AGCGGTCAAA
	451	CTCGACAATC	GGACCATTTA	CACGCGCTGC	GTATCCGCCA	AAGGCAATA
	501	CTACGCCACA	CCGCAAAAC	TGAACGCCGA	TTACCATTTT	GAGCAAAGTG
	551	TGCCTGCCGA	TATTTATTAC	ACGGTTACTG	AAGAACATAC	CGACAAATCC
50	601	AAGCTGTTTG	CAATATCTT	ATATACGCCC	CCCTTTTGA	TACTGGATGC
	651	GGCGGGCGCG	GTA	CTGGCCT	GGCTCTGGGT	GCGGTCGTGG
	701	ATGCCGCCCG	CAAATGA			

This corresponds to the amino acid sequence <SEQ ID 158; ORF28-1>:

55	1	MLFRKTTAAV	LAATLMLNGC	TLMLWGMNPN	VSETITRKHV	DKDQIRAFGV
	51	VAEDNAQLEK	GSLVMMGGKY	WFVVPEDSA	KLTGILKAGL	DKPFQIVEDT
	101	PSYARHQALP	VKLESPGSON	FSTEGLCRLY	DTDGPADIAK	LKQLGFEAVK
	151	LDNRTIYTRC	VSAKGKYYAT	PQKLNADYHF	EQSVADIYY	TVTEHTDKS

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF28 shows 79.2% identity over a 120aa overlap with an ORF (ORF28a) from strain A of *N.*

		10	20	30	40	50	60
	orf28.pep	<u>MLFRKTTAAVLAHTLMLNGCTLMLWGMNPNVSETITRKHVXKDQIRXFGVVVAEDNAQLEK</u>					
		::: : :					
10	orf28a	<u>MLFRKTTAAVLAATLMLNGCTVMWMGMSPFSETTARKHVDKDKQIRAFGVVVAEDNAQLEK</u>					
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf28.pep	GSLVMMGGKYWFVNPEDSAXXTGILXAGLDKFQIVXDTPSYXCHQALPVKLGSXG SQN					
		::: : :					
15	orf28a	GSLVMMGGKYWFVNPEDSAKLTGILKAGLDKQFQMVEPNRFA-YQALPVKLES PASQN					
		70	80	90	100	110	
	orf28a	FSTEGLC LRYDTRDPADI AKLKQ LEFEAVELDNRTIYTRCVS A SKGY AT POKLNAD YHF					
		120	130	140	150	160	170

	1	ATGTTGTTCC	GTA AACGAC	CGCCGCGGTT	TTGGCGGCAA	CCTTGATGTT
	51	GAACGGCTGT	ACGGTAATGA	TGTGGGGTAT	GAACAGCCCG	TTCAGCGAAA
	101	CGACGCCCCG	CAACACGTT	GACAAGGACC	AAATCCGCGC	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGA AAG	GGCAGCTGG	TGATGATGGG
25	201	CGGGAAATAC	TGGTTCGTCG	TCAATCCTGA	AGATTCGGCG	AAGCTGACGG
	251	GCATTTTGAA	GGCCGGGTTG	GACAAGCAGT	TTCAAATGGT	TGAGCCCAAC
	301	CCGCGCTTTG	CTACCAAGC	CCTGCCGGTC	AAACTCGAAT	CGCCCGCCAG
	351	CCAGAATTTC	AGTACCGAAG	GCCTTTGCCT	GCGCTACGAT	ACCGACAGAC
	401	CTGCCGACAT	CGCCAAGCTG	AAACAGCTTG	AGTTTGAAGC	GGTCGAACTC
30	451	GACAATCGGA	CCATTTACAC	GCGCTGCGTC	TCGCCCAAAG	GCAAATACTA
	501	CGCCACACCG	CAAAAAC TGA	ACGCCGATTA	TCATTTTGAG	CAAAGTGTGC
	551	CTGCCGATAT	TTATTACACG	GTTACGAAAA	AACATACCGA	CAAATCCAAG
	601	TTGTTTGAAA	ATATTGCATA	TACGCCCAAC	ACGTTGATAC	TGGATGCGGT
	651	GGGCGCGGTG	CTGGCCTTGC	CTGTGCGGCG	GTTGATTGCA	GCCACGAATT
35	701	CCTCAGACAA	ATGA			

40

1	MLFRKTTAAV	LAATLMLNGC	TVMWGMNSP	FSETTARKHV	DKDQIRAFGV
51	VAEDNAQLEK	GSLVMGGKY	WVVPNPDSA	KLTGILKAGL	DKQFQMVEPN
101	PRFAYQALPV	KLESPASQNF	STEGLCLRYD	TDRPADIACL	KQLEFEAVEL
151	DNRTIYTRCV	SAKGYYATP	QKLNADYHFE	QSVPADIIYTT	VTKKHTDKSK
201	LFENIAYTPT	TLILDVAGAV	LALPVAALIA	ATNSSDK*	

		10	20	30	40	50	60	
45	orf28a.pep	MLFRKTTAAVLAATLMLNGCTVMMWGMNSPFSETTARKHVDKDKQIRAFGVVAEDNAQLEK						
	orf28-1	MLFRKTTAAVLAATLMLNGCTLMLWGMNPNVSETITRKHVDKDKQIRAFGVVAEDNAQLEK						
		10	20	30	40	50	60	
		70	80	90	100	110	119	
50	orf28a.pep	GSLVMMGGKYWFVVPNPEDSAKLTGILKAGLDKQFQMVPEPNRFA-YQALPVKLESPASQN						
	orf28-1	GSLVMMGGKYWFVVPNPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSQN						
		70	80	90	100	110	120	
55		120	130	140	150	160	170	179
	orf28a.pep	FSTEGLCRLRYDTRDPADI AKLKQLE FEAVELDNRTIYTRCVSAKGKYYATPQKLNADYHF						
	orf28-1	FSTEGLCRLRYD TKPADI AKLKQLGFEAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF						
		130	140	150	160	170	180	

ORF28 shows 84.2% identity over a 120aa overlap with a predicted ORF (ORF28.ng) from *N.*

orf28.pep	MLFRKTTAAVLAHTLMLNGCTLMWGMNNPVSETITRKHVXKDQIRXFGVVAEDNAQLEK	60
	: : : : :	
orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSQTITRKHVDKDIARAFGVVAEDNAQLEK	60
orf28.pep	GSLVMGGKYWFVNPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGSXG SQN	120
	: : : : :	
orf28ng	GSLVMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPG SQN	120

20	1	ATGTTGTTCC	GTA AACGAC	CGCCGCCGTT	TTGGCGGC AA	CCTTGATACT
	51	GAACGGCTGT	ACGATGATGT	TGCGGGGGAT	GAACAACCCG	GTCAGCCAAA
	101	CAATCACCCG	CAAAACACGTT	GACAAAGACC	AAATCCGCGC	CTTCGGTGTG
	151	GTTGCCGAAG	ACAATGCCCA	ATTGGA AAAG	GGCAGCGCTG	TGATGATGGG
	201	CGGGA AATAC	TGGTTCCCG	TCAATCCCGA	AGATTCGGCG	AAGCTGACGG
25	251	GCCTTTTGAA	GGCCGGGTTG	GACAAGCCCT	TCCAAATAGT	TGAGGATACC
	301	CCGAGCTATG	CCCGCCACCA	AGCCCTGCGC	GCTCAATTCG	AAGCGCCCGG
	351	CAGCCAGAT	TTCAGTACCG	GAGGCTTTTG	CCTCGCGTAT	GATACCGGCA
	401	GACCTGACGA	CATCGCCAAG	CTGAACACAG	TTGAGTTTAA	AGCGGTCAAA
	451	CTCGACAATC	GGACCAATTA	CACGCGCTGC	GTATCCGCCA	AAGGCAATA
30	501	CTACGCCACG	CCGCAAAAAC	TGAACGCCGA	TTATCATTTT	GACAAAGTG
	551	TGCCCGCCGA	TATTTATTAT	ACGGTTACTG	AAAAACATAC	CGACAAATCC
	601	AAGCTGTTTG	GAAATATCTT	ATATACGCCC	CCCTTGTTGA	TATTGGATGC
	651	GGCGCGCCGC	GTGCTGGTCT	TGCCTATGGC	TCTGATTGCA	GCCGCGAATT
	701	CCTCAGACAA	ATGA			

35

1	<u>MLFRKTTAAV</u>	<u>LAATLILNGC</u>	<u>TMMLRGMNPN</u>	<u>VSQITIRKHW</u>	<u>DKDQIRAFGV</u>
51	<u>VAEDNAQLEK</u>	<u>GSLVMGGGKY</u>	<u>WFAVNPEDSA</u>	<u>KLTGLLKAGL</u>	<u>DKFPQIVEDT</u>
101	<u>PSYARHQALP</u>	<u>VKFEAPGSON</u>	<u>FSTGGGLCLRY</u>	<u>DTGRPPDIAK</u>	<u>LKQLEFKAVK</u>
151	<u>LDNRTIYTRC</u>	<u>VSAGKYYAT</u>	<u>PQKLNADYHF</u>	<u>EQSVPADIIY</u>	<u>TVTEKHTDKS</u>
201	<u>KLFGNILYTP</u>	<u>PLLILDAAA</u>	<u>VLVLPMALIA</u>	<u>AANSSDK*</u>	

		10	20	30	40	50	60
	orf28-1.pep	MLFRKTTAAVLAATLMLNGCTLMLWGMNPNVSETITRKHVDKDQIRAFGVVAEDNAQLEK					
	orf28ng	MLFRKTTAAVLAATLILNGCTMMLRGMNPNVSQTITRKHVDKDQIRAFGVVAEDNAQLEK					
45		10	20	30	40	50	60
	orf28-1.pep	GSLVMMGGKYWFVVPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSQN					
	orf28ng	GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSQN					
50		70	80	90	100	110	120
	orf28-1.pep	GSLVMMGGKYWFVVPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSQN					
	orf28ng	GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSQN					
		70	80	90	100	110	120
	orf28-1.pep	FSTEGCLCRYDTPDKPADI AKLKQLGF EAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF					
	orf28ng	FSTGGLCRYDTPGRPDDIAKLKQLEFKAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF					
55		130	140	150	160	170	180
	orf28-1.pep	FSTEGCLCRYDTPDKPADI AKLKQLGF EAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF					
	orf28ng	FSTGGLCRYDTPGRPDDIAKLKQLEFKAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF					
		130	140	150	160	170	180
	orf28-1.pep	EQSVPADIYYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX					
	orf28ng	EQSVPADIYYTVTEKHTDKSKLFGNILYTPPLILDAAAAVLVLPALIAAANSSDKX					
60		190	200	210	220	230	239
	orf28-1.pep	EQSVPADIYYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX					
	orf28ng	EQSVPADIYYTVTEKHTDKSKLFGNILYTPPLILDAAAAVLVLPALIAAANSSDKX					

190 200 210 220 230

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF28-1 (24kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 6A shows the results of affinity purification of the GST-fusion protein, and Figure 6B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result. These experiments confirm that ORF28-1 is a surface-exposed protein, and that it may be a useful immunogen.

Example 20

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 163>:

```
15      1  ..GTCAGTCCTG TACTGCCTAT TACACACGAA CGGACAGGGT TTGAAGGTGT
      51  TATCGGTAT  GAAACCCATT TTTCAGGGCA CGGACATGAA GTACACAGTC
      101  CGTTCGATCA TCATGATTCA AAAAGCACTT CTGATTTCAG CGGCGGTGTA
      151  GACGGCGGTT TTACTGTTTA CCAACTTCAT CGAACATGGT CGGAAATCCA
      201  TCCGGAGGAT GAATATGACG GGCCGCAAGC AGCG.ATTAT CCGCCCCCCG
      251  GAGGAGCAAG GGATATATAC AGCTATTATG TCAAAGGAAC TTCAACAAAA
      301  ACAAAGACTA GTATTGTCCC TCAAGCCCCA TTTTCAGACC GTTGGCTAGA
      351  AGAAAATGCC GGTGCCGCCT CTGGT..
```

This corresponds to the amino acid sequence <SEQ ID 164; ORF29>:

```
25      1  ..VSPVLPITHE RTGFEGVIGY ETHFSGHGHE VHSPFDHHDS KSTSDFSGGV
      51  DGGFTVYQLH RTWSEIHPED EYDGPQAAXY PPPGGARDIY SYVVKGTSTK
      101  TKTSIVPQAP FSDRWLEENA GAASG..
```

Further work revealed the complete nucleotide sequence <SEQ ID 165>:

```
30      1  ATGAATTGCG CTATTCAAAA ATTCATGATG CTGTTTGCGAG CAGCAATATC
      51  GTTGCTGCAA ATCCCCATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
      101  GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG
      151  TTTGGTAATG CTCGCGGCAG TGTTAAAAAG CGGGTTTACG CCGTCCAGAC
      201  ATTTGATGCA ACTGCGGTCA GTCCTGTACT GCCTATTACA CACGAACGGA
      251  CAGGGTTTGA AGGTGTTATC GGTATGAAA CCCATTTTTT AGGGCACGGA
      301  CATGAAGTAC ACAGTCCGTT CGATCATCAT GATTCAAAAA GCACTTCTGA
      351  TTTAGCGGCG GGTGTAGACG GCGGTTTAC TGTTTACCAA CTTTCATCGAA
      401  CAGGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC
      451  GATTATCCGC CCCCCGAGG AGCAAGGGAT ATATACAGCT ATTATGTCAA
      501  AGGAACCTCA ACAAACAACA AACTAATAT TGTCCCTCAA GCCCATTTT
      551  CAGACCGTTG GCTAAAAGAA AATGCCGGTG CCGCCTCTGG TTTTTCAGC
      601  CGTGCGGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA
      651  TTGGTGGGCT AACCGTATGG ATGATGTTTG CCGCATCGTC CAAGGTGCGG
      701  TTAATCCTTT TTAATGGGT TTTCAAGGAG TAGGGATTGG GGCAATTACA
      751  GACAGTGCGA TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA
      801  AGGTATTAAT GATTTAGGAA AATTAAGTCC GGAAGCACAA CTGCTGCCG
      851  CGAGCCTATT ACAGGACAGT GCTTTTGCGG TAAAAGACGG TATCAACTCT
      901  GCCAAACAAT GGGCTGATGC CCATCCAAAT ATAACAGCTA CTGCCCAAAC
      951  TGCCCTTTCC GCAGCAGAGG CCGCAGGTAC GGTGTGAGA GGTAAAAAAG
      1001  TAGAACTTAA CCGACTAAA TGGGATTGGG TAAAAATAC CGGTTATAAA
      1051  AAACCTGCTG CCCGCCATAT GCAGACTTTA GATGGGGAGA TGGCAGGTGG
      1101  GAATAAACCT ATTAATCTT TACCAACAG TGCCGCTGAA AAAAGAAAAC
      1151  AAAATTTTGA GAAGTTTAA AGTAAGTGA GTTCAGCAAG TTTTGATTCA
```

-144-

1201 GTGCACAAAA CACTAACTCC CAATGCACCT GGTATTTTAA GTCCTGATAA
 1251 AGTTAAAACT CGATACACTA GTTAGATGG AAAAATTACA ATTATAAAAG
 1301 ATAACGAAAA CAACTATTTT AGAATCCATG ATAATTCAG AAAACAGTAT
 1351 CTTGATTCAA ATGGTAATGC TGTGAAAACC GGTAAATTTAC AAGGTAAGCA
 1401 AGCAAAAGAT TATTACAAC AACAACTCA TATCAGGAAC TTAGACAAAT
 1451 GA

This corresponds to the amino acid sequence <SEQ ID 166; ORF29-1>:

1 MNLPQKFM LFAAAISLLQ IPISHANGLD ARLRDMQAK HYPEGGKYHL
 51 FGNARGSVKK RYVAVQTFDA TAVSPVLPIT HERTGFEGVI GYETHFSGHG
 101 HEVHSPFDHH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS
 151 DYPPPGGARD IYSYVKGTS TKTKTNIVPQ APFSDRWLKE NAGAASGFFS
 201 RADEAGKLIW ESDPNKNWWA NRMDDVRGIV QGAVNPFLMG FQGVGIGAIT
 251 DSAVSPVTD TAAQTLQGIN DLGKLSPEAQ LAAASLLQDS AFAVKDGIN
 301 AKQWADAHFN ITATAQTALS AAEAAGTVWR GKKVELNPTK WDWVKNTRYK
 15 351 KPAARHMQTL DGEMAGGNKP IKSLPNSAAE KRKQNFKEFN SNWSSASFDS
 401 VHKTLPNAP GILSPDKVKT RYTSLDGKIT IIKDNENNYF RIHDNSRKQY
 451 LDSNGNAVKT GNLQKQAKD YLQQQTHIRN LDK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF29 shows 88.0% identity over a 125aa overlap with an ORF (ORF29a) from strain A of *N. meningitidis*:

				10	20	30
	orf29.pep			VSPVLPIT	HERTGFEGVI	GYETHFSGHG
				10	20	30
25	orf29a	EPGGKYHL	FGNARGSVKNRYVAVQTFD	ATAVGPI	LPI	HERTGFEGVI
		50	60	70	80	90
		40	50	60	70	80
30	orf29.pep	VHSPFDH	DSKSTSDFS	GGVDGGFTVYQLHRTWSEIHP	EDGYDGPQAAXYPPPGGARDIY	
	orf29a	VHSPFDH	DSKSTSDFS	GGVDGGFTVYQLHRTWSEIHP	EDGYDGPQGS	SDYPPPGGARDIY
		110	120	130	140	150
35	orf29.pep	SYVVKGTSTKTKTSIVPQAPFSDRWLEENAGAASG				
	orf29a	XXYVVKGTSTKTKSNIVPRAPFSDRWLEENAGAASGFFSR	ADEAGKLIW	ESDPNKNWWANR		
		170	180	190	200	210
40	orf29a	MDDIRGIVQ	GAVNPF	LMGFQGVGIGAIT	DSAVSPVTD	TAAQTLQGXNHLGXLSPEAQLA
		230	240	250	260	270

The complete length ORF29a nucleotide sequence <SEQ ID 167> is:

1 ATGAATTNGC CTATTCAAAA ATTCATGATG CTGTTTGAC CAGCAATATC
 45 51 GTNGCTGCAA ATCCCNATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
 101 GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG
 151 TTTGGTAATG CTCGCGGACG TGTTAAAAAT CGGGTTTACG CCGTCCAAAC
 201 ATTTGATGCA ACTGCGGTCTG GCCCATACT GCCTATTACA CACGAACGGA
 251 CAGGATTGA AGGCATTATC GGTATGAAA CCCATTTTTC AGGACATGGA
 301 CATGAAGTAC ACAGTCCGTT CGATAATCAT GATTCAAAA GCACTTCTGA
 50 351 TTTCAGCGGC GGCCTAGACG GTGGTTTAC CGTTTACCA CTTTCATCGGA
 401 CAGGGTCCGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC
 451 GATTATCCG CCCCAGGAGG AGCAAGGGAT ATATACANNT ANTATGTCAA
 501 AGGAACTTCA ACAAACAA AGAGTAATAT TGTTCCCGCA GCCCATTTT
 55 551 CAGACCGCTG GCTAAAAGAA AATGCCGGTG CCGCCTCTGG TTTTTCAGC
 601 CGTGCTGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA
 651 TTGGTGGGCT AACCGTATGG ATGATATTCG CGGCATCGTC CAAGGTGCGG
 701 TTAATCCTTT TTAATGGGT TTTCAAGGAG TAGGGATTGG GGCATTACA
 751 GACAGTGCAG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA
 801 AGGTATNAAT CATTTAGGAA ANTAAAGTCC CGAAGCACAA CTTGCGGCTG
 60 851 CAACCGCATT ACAAGACAGT GCTTTGCGG TAAAGACGG TATCAATTCC
 901 GCCAGACAAT GGGCTGATGC CCATCCGAAT ATAAGTCAA CAGCCCAAC

951 TGCCCTTGCC GTAGCAGANG CCGCAACTAC GGTGTGGGGC GGTAAAAAAG
 1001 TAGAACTTAA CCCGACCAAA TGGGATTGGG TTAATAATAC NGGCTATAAN
 1051 ACACCTGCTG TTCGCACCAT GCATACTTG GATGGGGAAA TGGCCGGTGG
 1101 GAATAGACCG CCTAAATCTA TAACGTCCAA CAGCAAAGCA GATGCTTCCA
 1151 CACAACCGTC TTTACAAGCG CAACTAATG GAGAACAAT TANNNNNGGG
 1201 CATGCTTATA ACAAGCATGT CATAAGACAA CAAGAATTA CCGATTTAAA
 1251 TATCAATTCA CCAGCAGATT TTGCTCGGCA TATTGAAAAT ATTGTTAGCC
 1301 ATCCANCAAA TATGAAAGAG TTACCTCGCG GTAGAACTGC GTATTGGGAT
 1351 NATAAAACAG GGACNATAGT TATCCGAGAT AAAAATTCTG ACGATGGAGG
 1401 TACAGCATT AGACCAACAT CAGGTAAAAA ATATTATGAT GATTATAG

This encodes a protein having amino acid sequence <SEQ ID 168>:

1 MNXPIQKFM LFAAAISXLO IPISHANGLD ARLRDDMQAK HYEPPGKYHL
 51 FGNARGSVKN RVYAVQTFDA TAVGPILPIT HERTGFEGII GYETHFSGHG
 101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS
 151 DYPPPGGARD IYXXYVKGTS TKTKSNIVPR APFSDRWLKE NAGAASGFFS
 201 RADEAGKLIW ESDPNKNWWA NRMDDIRGIV QGAVNPFLMG FQGVGIGAIT
 251 DSAVSPVTD TAAQTLQGXN HLGXLSPEAQ LAAATALQDS AFAVKDGINS
 301 ARQWADAHFN ITATAQTALA VAXAATTVWG GKKVELNPTK WDWVKNTGYX
 351 TPAVRTMHTL DGEMAGGNRP PKSITSNSKA DASTQPSLQA QLIGEQIXXG
 401 HAYNKHVIRQ QEFTDLNINS PADFARHIEN IVSHPNXMK E LPRGRTAYWD
 451 XKTGTIVIRD KNSDDGGTAF RPTSGKKYYD DL*

ORF29a and ORF29-1 show 90.1% identity in 385 aa overlap:

		10	20	30	40	50	60
25	orf29a.pep	MNXPIQKFMMLFAAAISXLOIPISHANGLDARLRDDMQAKHYEPPGKYHLFGNARGSVKN					
	orf29-1	MNLPIQKFMMLFAAAISLLOIPISHANGLDARLRDDMQAKHYEPPGKYHLFGNARGSVKK					
		10	20	30	40	50	60
30	orf29a.pep	RVYAVQTFDATAVGPILPITHERTGFEGIIIGYETHFSGHGHEVHSPFDNHDSKSTSDFSG					
	orf29-1	RVYAVQTFDATAVSPVLPITHERTGFEGVIGYETHFSGHGHEVHSPFDHDSKSTSDFSG					
		70	80	90	100	110	120
35	orf29a.pep	GVDGGFTVYQLHRTGSEIHPEDGYDGPQGS DYPPPGGARDIYXXYVKGTS TKTKSNIVPR					
	orf29-1	GVDGGFTVYQLHRTGSEIHPEDGYDGPQGS DYPPPGGARDIYSYVKGTS TKTKTNIVPQ					
		130	140	150	160	170	180
40	orf29a.pep	APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMDDIRGIVQGAVNPFLMG					
	orf29-1	APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMDDVRGIVQGAVNPFLMG					
45		190	200	210	220	230	240
50	orf29a.pep	FQGVGIGAITDSAVSPVTD TAAQTLQGXNHLGXLSPEAQ LAAATALQDS AFAVKDGINS					
	orf29-1	FQGVGIGAITDSAVSPVTD TAAQTLQGINDLGKLSPEAQ LAAASLLQDS AFAVKDGINS					
		250	260	270	280	290	300
55	orf29a.pep	ARQWADAHFNITATAQTALAVAXAATTVWGGKKVELNPTKWDVKNKGYXTPAVRTMHTL					
	orf29-1	AKQWADAHFNITATAQTALSAEAAAGTVWRGKKVELNPTKWDVKNKGYKKPAARHMQTL					
		310	320	330	340	350	360
60	orf29a.pep	DGEMAGGNRP PKSITSNSKADASTQPSLQAQLIGEQIXXGHAYNKHVIRQQEFTDLNINS					
	orf29-1	DGEMAGGNKPIKSLP-NSAAEKRKQNF EKFN SNWSSASFDSVHKTLTPNAPGILSPDKVK					
		370	380	390	400	410	420

Homology with a predicted ORF from *N.gonorrhoeae*

ORF29 shows 88.8% identity over a 125aa overlap with a predicted ORF (ORF29.ng) from *N.*

gonorrhoeae:

5	orf29.pep	VSPVLPITHERTGFEGVIGYETHFSGHGHE	30
	orf29ng	EPGGKYHLFGNARGSVKNRVCVQTFDATAVGPILPITHERTGFEGVIGYETHFSGHGHE	102
10	orf29.pep	VHSPFDHDSKSTSDFSGGVDGGFTVYQLHRTWSEIHPEDGYDGPQAAAXYPPPGGARDIY	90
	orf29ng	VHSPFDNHDSKSTSDFSGGVDGGFTVYQLHRTGSEIHPEDGYDGPQGGGYPPPGGARDIY	162
15	orf29.pep	SYVVKGTSTKTKTSIVPQAPFSDRWLEENAGAASG	125
	orf29ng	SYHIKGTSTKTKINTVPQAPFSDRWLEENAGAASGFLSRDEAGKLIWENDPKNWRANR	222

- 15 The complete length ORF29ng nucleotide sequence <SEQ ID 169> is predicted to encode a protein having amino acid sequence <SEQ ID 170>:

	1	MNLPIQKFM LFAAAISLLQ IPISHANGLD ARLRDDMQAK HYEPPGGKYHL
	51	FGNARGSVKN RVCVQTFDA TAVGPILPIT HERTGFEGVI GYETHFSGHG
	101	HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGG
20	151	GYPPPGGARD IYSYHIKGT TKTINTVPQ APFSDRWLKE NAGAASGFLS
	201	RADEAGKLIW ENDPDKNWR NRMDDIRGIV QGAVNPFLTQ FQGLGVGAI
	251	DSAVSPVTYA AARKTLQGIH NLGNLSPEAQ LAAATALQDS AFAVKDSINS
	301	ARQWADAHPN ITATAQTALA VTEAATTWVG GKKVELNPAK WDWKNTGYK
	351	KPAARHMQTV DGEMAGGNKP LESKNTVTN NFFENTGYE KVLQASNGD
25	401	YHGFQSVDA FSENGTVIQI VGGDNIVRHK LYIPGSYKKG DGNFEYIREA
	451	DGKINHRLFV PNQQLPEK*

In a second experiment, the following DNA sequence <SEQ ID 171> was identified:

	1	atgAATTTGC CTATTCAAAA ATTCATGATG ctgttggcAg cggcaatata
	51	gatgctGCat ATCCCCATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
30	101	GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGCAA ATACCATCTG
	151	TTTGGTAATG CTCGCGGCAG TGTTAAAAAT CGGGTTTGGC CCGTCCAAAC
	201	ATTTGATGCA ACTGCGGTCTG GCGGTCATCT GCCTATTACA CACGAACGGA
	251	CAGGATTTGA AGGTGTTATC GGCTATGAAA CCCATTTC AGGACACGGA
	301	CACGAAGTAC ACAGTCCGTT CGATAATCAT GATTCAAAAA GCACTTCTGA
35	351	TTTCAGCGGC GCGGTAGACG GCGGTTTTC CGTTTACCAA CTTTCATCGGA
	401	CAGGGTCGGA AATACATCCC GCAGACGGAT ATGACGGGCC TCAAGCGGGC
	451	GGTTATCCGG AACCAACAGG GGCAAGGGAT ATATACAGCT ACCATATCAA
	501	AGGAACCTCA ACCAAAAACA AGATAAACAC TGTTCCGCAA GCCCTTTT
	551	CAGACCGCTG GCTAAAAGAA AATGCCGGTG CCGCTTCCGG TTTTCTCAGC
40	601	CGTGCGGATG AAGCAGGAAA ACTGATATGG GAAAACGACC CCGATAAAAA
	651	TTGGCGGGCT AACCGTATGG ATGATATTCT CGGCATCTGC CAAGGTGCGG
	701	TTAATCCTTT TTTAACGGGT TTTCAAGGGG TAGGGATTGG GGCAATTACA
	751	GACAGTGGCG TAAGCCCGGT CACAGATACA GCCGCTCAGC AGACTCTACA
	801	AGGTATTAAAT GATTAGGAA ATTTAAGTCC GGAAGCACAA CTGCGCGCG
45	851	CGAGCCTATT ACAGGACAGT GCCTTTGCGG TAAAAGACGG CATCAATTCC
	901	GCCAGACAAT GGGCTGATGC CCATCCGAAT ATAACAGCAA CAGCCCAAAC
	951	TGCCCTTGCC GTAGCAGAGG CCGCAGGTAC GGTTTGGCGC GGTAAAAAAG
	1001	TAGAACTTAA CCCGACCAA TGGGATTGGG TAAAAAATAC CGGCTATAAA
	1051	AAACCTGCTG CCCGCCATAT GCAGACTGTA GATGGGGAGA TGGCAGGGGG
50	1101	GAATAGACCG CCTAAATCTA TAACGTGCGA AGGAAAAGCT AATGCTGCAA
	1151	CCTATCCTAA GTTGGTTAAT CAGCTAAATG AGCAAACTT AAATAACATT
	1201	GCGGCTCAAG ATCCAAGATT GAGTCTAGCT ATTCATGAGG GTAAAAAATA
	1251	TTTTCCAATA GGAAGTCAA CTTATGAAGA GGCAGATAGA CTAGGTAAAA
	1301	TTTGGGTTGG TGAGGGTGCA AGACAACTA GTGGAGGCGG ATGGTTAAGT
55	1351	AGAGATGGCA CTCGACAATA TCGGCCACCA ACAGAAAAAA AATCACAATT
	1401	TGCAACTACA GGTATTCAAG CAAATTTTGA AACTTATACT ATTGATTCAA
	1451	ATGAAAAAAG AAATAAAATT AAAAATGGAC ATTTAAATAT TAGGTAA

This encodes a protein having amino acid sequence <SEQ ID 172; ORF29ng-1>:

	1	MNLPIQKFM LLAAAISMLH IPISHANGLD ARLRDDMQAK HYEPPGGKYHL
60	51	FGNARGSVKN RVCVQTFDA TAVGPILPIT HERTGFEGVI GYETHFSGHG

-147-

101 HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP ADGYDGPQGG
 151 GYPEPQGARD IYSYHIKGTST TKTINTVPQ APFSDRWLKE NAGAASGFLS
 201 RADEAGKLIW ENDPDKNWR NRMDDIRGIV QGAVNPFLTQ FQGVGIGAIT
 251 DSAVSPVTD TAAQQTLOGIN DLGNLSPEAQ LAAASLLQDS AFAVKDGINS
 301 ARQWADAHPN ITATAQTALA VAEAAGTVWR GKKVELNPTK WDVKNTGYK
 351 KPAARHMOTV DGEMAGGNRP PKSITSEGKA NAATYPKLVN QLNEQNLNNI
 401 AAQDPRLSLA IHEGKKNFPI GTATYEEADR LGKIWVGEA RQTSGGGWLS
 451 RDGTRQYRPP TEKKSQFATT GIQANFETY IDSNEKRNI KNHGLNIR*

ORF29ng-1 and ORF29-1 show 86.0% identity in 401 aa overlap:

10	orf29ng-1.pep	10	20	30	40	50	60
	orf29-1	10	20	30	40	50	60
15	orf29ng-1.pep	70	80	90	100	110	120
	orf29-1	70	80	90	100	110	120
20	orf29ng-1.pep	130	140	150	160	170	180
	orf29-1	130	140	150	160	170	180
25	orf29ng-1.pep	190	200	210	220	230	240
	orf29-1	190	200	210	220	230	240
30	orf29ng-1.pep	250	260	270	280	290	300
	orf29-1	250	260	270	280	290	300
35	orf29ng-1.pep	310	320	330	340	350	360
	orf29-1	310	320	330	340	350	360
40	orf29ng-1.pep	370	380	390	400	410	419
	orf29-1	370	380	390	400	410	420
45	orf29ng-1.pep	420	430	440	450	460	479
	orf29-1	420	430	440	450	460	480

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 21

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 173>:

-148-

1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAATGTTCC
 101 ACACGCGGGC AGATGCACCG ATGCAG...

This corresponds to the amino acid sequence <SEQ ID 174; ORF30>:

5 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QMFHTRADAP MQ..

Further work revealed the complete nucleotide sequence <SEQ ID 175>:

1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC
 101 ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
 151 ATGAAGGAGA CAGAGGGGGC GTTCTTCCA TTGGCTATCT TGGGTGGTGC
 201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
 251 GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
 301 CCTGGTGGTG TAGGCGCCGC AGGAAAGGTT GTTCCCTTTG CTAAATATGG
 351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCTT TTCGGTAATA
 15 401 GAACAGGTCA TCCTATTGGA AAATTTCCCC ATTATCATCG TCGAGTTACG
 451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
 501 TTGGAATCA AAATCTACGG ACAGATCATG GAAAACCGC TTCTAA

This corresponds to the amino acid sequence <SEQ ID 176; ORF30-1>:

20 1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
 51 MKETEGAFLEP LAILGGAAIG MWTQHGFSA TTGRPASVRD VAIAGGLGAI
 101 PGGVGAAGKV VSFAYGREI KIGNNMRIAP FGNRTGHPIG KFPYHRRVT
 151 DNTGKTLPGQ GIGRHRPWES KSTDRSWKNR F*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF30 shows 97.6% identity over a 42aa overlap with an ORF (ORF30a) from strain A of *N. meningitidis*:

	10	20	30	40
orf30.pep	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ			
	:			
30 orf30a	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP			
	10	20	30	40 50 60
orf30a	LXILGGAAIGMWTQHGFSAATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI			
	70	80	90	100 110 120

35 The complete length ORF30a nucleotide sequence <SEQ ID 177> is:

1 ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51 CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC
 101 ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
 151 ATGAAGGANA CAGNGGGGGC GTTCTTCCA TTGGNTATCT TGGGTGGTGC
 40 201 TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
 251 GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
 301 CCTGGTGGTG TAGGCGCCGC AGGAAAGGTT GTTCCCTTTG CTAAATATGG
 351 ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCTT TTCGGTAATA
 401 GAACAGGTCA TCCTATTGGN AAATTTCCCC ATTATCATCG TCGAGTTACG
 45 451 GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
 501 TTGGAATCA AAATCTACGG ACAGATCATG GAAAACCGC TTCTAA

This encodes a protein having amino acid sequence <SEQ ID 178>:

1 MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
 51 MKXTXGAFLP LXILGGAAIG MWTQHGFSA TTGRPASVRD VAIAGGLGAI
 50 101 PGXVGAAGKV VSFAYGREI KIGNNMRIAP FGNRTGHPIG KFPYHRRVT
 151 DNTGKTLPGQ GIGRHRPWES KSTDRSWKNR F*

ORF30a and ORF30-1 show 97.8% identity in 181 aa overlap:

orf30a.pep MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP 60

	orf30-1	 MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP	60
5	orf30a.pep	LXILGGAAIGMWTQHGFYSYATTGRPASVRDVAIAGGLGAI PGXVGAAGKVVSFAKYGREI	120
	orf30-1	 LAILGGAAIGMWTQHGFYSYATTGRPASVRDVAIAGGLGAI PGGVGAAGKVVSFAKYGREI	120
	orf30a.pep	KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR	180
10	orf30-1	 KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR	180
	orf30a.pep	FX	
15	orf30-1	FX	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF30 shows 97.6% identity over a 42aa overlap with a predicted ORF (ORF30.ng) from *N. gonorrhoeae*:

	orf30.pep	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ	42
20			
	orf30ng	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP	60

The complete length ORF30ng nucleotide sequence <SEQ ID 179> is

	1	ATGAAAAAAC	AAATCACC	AGCCGTAATG	ATGCTGTCTA	TGATCGCCCC
	51	CGCAATGGCA	AACGGATTGG	ACAATCAGGC	ATTGAAGAC	CAAGTGTTCC
25	101	ACACGCGGGC	AGATGCGCCG	ATGCAGTTGG	CGGAGCTTTC	TCAGAAGGAG
	151	ATGAAGGAGA	CTGAAGGGGC	TTTTCTTCCA	TTGGCTATCT	TGGGTGGTGC
	201	TGCCATTGGT	ATGTGGACAC	AGCATGGTTT	TAGTTATGCA	ACGACAGGCA
	251	GACCAGCTTC	TGTTAGAGAT	GTTGCTGGCG	GATTAGGCGC	AATTCCTGGT
	301	GATGTAGGTG	CTGCAGGAAA	GGTTGTTTCC	TTTGCTAAAT	ATGGACGTGA
30	351	GATTAAAATC	GGCAATAATA	TGCGGATAGC	CCCTTTCGGT	AATAGAACAG
	401	GTCATCCTAT	TGGAAATTTT	CCCCATTATC	ATCGTCGAGT	TACGGATAAT
	451	ACGGGCAAGA	CTTTGCCTGG	ACAGGGAATT	GGTCGTCATC	GCCCTTGGGA
	501	ATCAAAATCT	ACGGACAGAT	CATGAAAAAA	CCGCTTCTAA	

This encodes a protein having amino acid sequence <SEQ ID 180>:

35	1	MKKQITAAVM	MLSMIAPAMA	NGLDNQAFED	QVFHTRADAP	MQLAELSQKE
	51	MKETEGAFLP	LAILGGAAIG	MWTQHGFSA	TTGRPASVRD	VAGGLGAIPG
	101	DVGAAGKVVS	FAKYGREIKI	GNNMRIAPFG	NRTGHPIGKF	PHYHRRVTDN
	151	TGKTLPGQGI	GRHRPWESKS	TDRSWKNRF*		

ORF30ng and ORF30-1 show 98.3% identity in 181 aa overlap:

40		10	20	30	40	50	60
	orf30ng.pep	MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP					
	orf30-1						
		10	20	30	40	50	60
45		70	80	90	100	110	
	orf30ng.pep	LAILGGAAIGMWTQHGFYSYATTGRPASVRDVA--GGLGAIPGDVGAAGKVVSFAKYGREI					
	orf30-1						
50		70	80	90	100	110	120
		LAILGGAAIGMWTQHGFYSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI					
		120	130	140	150	160	170
	orf30ng.pep	KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR					
	orf30-1						
55		130	140	150	160	170	180
		KIGNNMRIAPFGNRTGHPIGKFPYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR					
		180					
	orf30ng.pep	FX					
60							
	orf30-1	FX					

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 181>:

This corresponds to the amino acid sequence <SEQ ID 182; ORF31>:

15 Further work revealed a further partial nucleotide sequence <SEQ ID 183>:

This corresponds to the amino acid sequence <SEQ ID 184; ORF31-1>:

Computer analysis of this amino acid sequence gave the following results:

ORF31 shows 76.2% identity over a 84aa overlap with a predicted ORF (ORF31.ng) from *N. gonorrhoeae*:

35 The complete length ORF31ng nucleotide sequence <SEQ ID 185> is:

BNSDOCID: <WO__9924578A2_1_>

701 GTTTGGATGC CCGTGATACC GATTTCACAC GTATTCTTGT ATGCCAACAA
751 AATCACCTTG ATCAGTACGG CCGAACAAAGC AGGCATTCGT AA

This encodes a protein having amino acid sequence <SEQ ID 186>:

5 1 MNKTLRVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIPTH
 51 SKAFCFSALG FSLCLALGTV NIAFADGIIT DKAAPKTOQA TILOTGNGIP
 101 QVNIQTPTSA GVSVNQYQAF DVGNRGAILN NSRSNTQTQL GGWIQGNPWL
 151 TRGEARVVVN QINSSHPSQL NGYIEVGRR AEVVIANPAG IAVNGGGFIN
 201 ASRATLTGQ PQYQAGDFSG FKIRQGNVAV AGHGLDARDT DFTRILVCQQ
 251 NHLDQYGRTS RHS*

10 This gonococcal protein shares 50% identity over a 149aa overlap with the pore-forming hemolysins-like HecA protein from *Erwinia chrysanthemi* (accession number L39897):

 orf31ng 96 GNGIQVNIQTPTSAGVSVNQAQFDVGNRGAILNNSRSN-TQTQLGGWIQGNPWLTRGE 154
 GNG+P VNI TP ++G+S N+Y F+V NRG ILNN + T +QLGG IQ NP L
 HecA 45 GNGVPVVNIATPDASGLSHNRYHDFVNDNRGLILNNGTARLTSPQLGGLIQNNPNLNGRA 104
15 Orf31ng 155 ARVVVNQINSSHPSQLNGYIEVGRRRAEVVIANPAGIAVNGGGFINASRATLTGQPPQY 214
 A ++N++ S + S+L GY+EV G+ A VV+ANP GI +G GF+N R TLTTG PQ+
 HecA 105 AAAILNEVVS PNRSRLAGYLEVAGQAANVVVANPYGITCSGCGFLNTPRLTLTTGTQPD 164
20 Orf31ng 215 -AGDFSGFKIRQGNVAVAGHGLDARDTDF 242
 AG SG +R G+ +I G GLDA +D+
 HecA 165 AAGGLSGLDVRGGDILIDGAGLDASRDY 193

Furthermore, ORF31ng and ORF31-1 show 79.5% identity in 83 aa overlap:

25 orf31-1.pep 10 20 30 40 50 60
 MNKTLRVIFNRKRGAVVAVAEETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCRSNIFS
 |||||
 orf31ng MNKTLRVIFNRKRGAVVAVAEETTKREGKSCADSGSGSVYVKSVSFIPTH-----SKAFC
 10 20 30 40 50
30 orf31-1.pep 70 80
 FSLGFSCLAVGTANIAFADGI
 || |||||:|:|:|
 orf31ng FSALGFSCLALGTVNIAFADGIITDKAAPKTOQATILQTGNGIQVNIQTPTSAGVSVN
 60 70 80 90 100 110

35 On this basis, including the homology with hemolysins, and also with adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 23

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 187>:

40 1 ATGAATACTC CTCCTTTTGT CTGTTGGATT TTTTGCAAGG TCATCGACAA
 51 TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGT TTGCACCGCG
 101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCCT
 151 GCGCTTTGCC CTGATTGCC CGATGTTCC TCGTTCATC AGGATATTCA
 201 TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCG..

45 This corresponds to the amino acid sequence <SEQ ID 188; ORF32>:

 1 MNTPPFVCWI FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
 51 ALCPDLFDPV CVHQDIHVRT WHSDAADIDT A..

Further work revealed the complete nucleotide sequence <SEQ ID 189>:

50 1 ATGAATACTC CTCCTTTTGT CTGTTGGATT TTTTGCAAGG TCATCGACAA
 51 TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGT TTGCACCGCG
 101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCCT

-152-

5
10
15
20

```

151 GCGCTTTGCC CTGATTGGCC CGATGTTCCC TCGGTTTCATC AGGATATTCA
201 TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTT
251 CCGATGTCGT CATCGAAACT TTTGCCCTGCG ACCTGCCCGA AAATGTGCTG
301 CACATTATCC GCCGACACAA GCCGCTTTGG CTGAATTGGG AATATTGAG
351 CCGGAGGAA AGCAATGAAA GGCTGCATCT GATGCCTTCG CCGCAGGAGG
401 GTGTTCAAAA ATATTTTGG TTTATGGGTT TCAGCGAAAA AAGCGGCGGG
451 TTGATACGCG AACGTGATTA CTGCGAAGCC GTCCGTTTCG ATACTGAAGC
501 CCGGAGAG CGGCTGATGC TGCCCGAAAA AAACGCCTCC GAATGGCTGC
551 TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
601 CAGGCAGGCA GCCCGATGAC ACTGTTGCTG GCGGGGACGC AAATCATCGA
651 CAGCCTCAAA CAAAGCGGCG TTATTCGCA AGATGCCCTG CAAAACGACG
701 GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
751 CCGCAACAGG ACTTCGACCA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT
801 CCGCGGCGAA GACAGTTTCG TCGCGGCCCA GCTTGGGGC AAACCTTCT
851 TTTGGCACAT CTACCCGCAA GACGAGAATG TCCATCTCGA CAAACTCCAC
901 GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCGAAA CCGTGTCCGC
951 ACACCGCCGT CTTTCGGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
1001 CACAACGCCT CGAATGTTGG CAAACCTGC AACAAATCA AACCGGTGG
1051 CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTCGGGC AGCCGTCAGC
1101 TCCTGAAAAA CTCGCTGCCT TTGTTTCAA GCATCAAAA ATACGCTAG

```

This corresponds to the amino acid sequence <SEQ ID 190; ORF32-1>:

25

```

1 MNTPPFVCWI FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
51 ALCPLDLPDVP CVHQDIHVRT WHSDAADIDT APVPDVVIET FACDLPENVL
101 HIIRRHKPLW LNWEYLSAEE SNERLHLMPS PQEGVQKYFW FMGFSEKSGG
151 LIRERDYCEA VRFDTEALRE RLMLPEKNAS EWLLFGYRSD VWAKWLEMWR
201 QAGSPMTLLL AGTQIIDSLE QSGVLPQDAL QNDGDVFQTA SVRLVKIPFV
251 PQQDFDQLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPQ DENVHLDKLH
301 AFWDKAHGFY TPETVSAHRR LSDDLNGGEA LSATORLECW QTLQHQHNGW
351 RQGAEDWSRY LFGQPSAPEK LAAFVSKHQK IR*w

```

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF32 shows 93.8% identity over a 81aa overlap with an ORF (ORF32a) from strain A of *N.*

meningitidis:

35

```

      10      20      30      40      50      60
orf32.pep MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVP
      |||||  |||||||  |||||||  |||||||  |||||||  |||||||
orf32a    MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX
      10      20      30      40      50      60

      70      80
orf32.pep CVHQDIHVRTWHSDAADIDTA
      |||||  |||||
orf32a    CVHQDIHVRTWHSDAADIDTAPVXDVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX
      70      80      90      100     110     120

```

45 The complete length ORF32a nucleotide sequence <SEQ ID 191> is:

50
55
60

```

1 ATGAATACTC CTCCTTTTTC TGCTGGANTT TTTTGCAAGG TCATCGACAA
51 TTTCCGCGAC ATCGGCGTTT CGTGGCGGCT TGCCCGTGT TGCACCGCG
101 AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
151 GCGCTTTGCC CTGATTGGCC CGATGTTTNC TCGGTTTCATC AGGATATTCA
201 TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTT
251 NCATGTCGT CATCGAAACT TTTGCCCTGCG ACCTGCCCGA AAATGTGCTG
301 CACATCATCC GCCGACACAA GCCGCTTTGG CTGAANTGGG AATATTGAG
351 CCGGAGGAN AGCAATGAAA GGCTGCACNT GATGCCTTCG CCGCAGGAGA
401 GTGTTCAAAA ATANTTTTGG TTTATGGGTT TCAGCGAANN NAGCGGCGGA
451 CTGATACGCG AACGCGATTA CTGCGAAGCC GTCCGTTTCG ATAGCGGAGC
501 CTTGCGCAAG AGGCTGATGC TTCCCGAAAA AAACGNCCCC GAATGGCTGC
551 TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
601 CAGGCAGGCA GTCCGTTGAC ACTTTTGCTG GCNCGGGGCG ANATTATCGA
651 CAGCCTCAAA CAAAACGGCG TTATTCGCA AGATGCCCTG CAAAACGACG
701 GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
751 CCGCAACAGG ACTTCGACAA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT

```

801 CCGCGGCGAA GACAGTTTCG TCGCGGCCCA GCTTGGCGGC AAACCTTCT
 851 TTTGGCACAT CTACCCGCAA GATGAGAATG TCCATCTCGA CAAACTCCAC
 901 GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGCATCGGC
 951 ACACCGCCGC CTTTCAGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
 5 1001 CACAACGCCT CGAATGTTGG CAAATCCTGC AACAACATCA AAACGGCTGG
 1051 CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTTGGGC AGCCTTCCGC
 1101 ATCCGAAAAA CTCGCCGCTT TTGTTTCAAA GCATCAAAA ATACGCTAG

This encodes a protein having amino acid sequence <SEQ ID 192>:

1 MNTPPFSAGX FCKVIDNFGD IGVSRLARV LHRELGWQVH LWTDDVSALR
 10 51 ALCPDLPDVX CVHQDIHVRT WHSDAADIDT APVXDVIET FACDLPENVL
 101 HIIRRHKPLW LXWEYLSAEX SNERLHXMPS PQESVXKXFW FMGFSEXSGG
 151 LIRERDYCEA VRFDSGALRK RLMLPEKNXP EWLLFGYRSD VWAKWLEMR
 201 QAGSPLTLL AGAXIIDLK QNGVIPQDAL QNDGDVFQTA SVRLVKIPFV
 251 PQQDFDKLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPO DENVHLDKLH
 15 301 AFWDKAHGFY TPETASAHRR LSDDLNGGEA LSATQRLCW QILQQHONGW
 351 RQGAEDWSRY LFGQPSASEK LAAFVSKHQK IR*

ORF32a and ORF32-1 show 93.2% identity in 382 aa overlap:

		10	20	30	40	50	60
20	orf32-1.pep	MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX					
	orf32a	MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX					
		10	20	30	40	50	60
		70	80	90	100	110	120
25	orf32-1.pep	CVHQDIHVRTWHSDAADIDTAPVPDVVIETFACDLPENVLHIIRRHKPLWLNWEYLSAEE					
	orf32a	CVHQDIHVRTWHSDAADIDTAPVXDVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX					
		70	80	90	100	110	120
		130	140	150	160	170	180
30	orf32-1.pep	SNERLHLMPSPOEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRERLMLPEKNAS					
	orf32a	SNERLHXMPSPOESVXKXFWFMGFSEXSGGLIRERDYCEAVRFDGALRKRMLPEKNXP					
		130	140	150	160	170	180
35		190	200	210	220	230	240
	orf32-1.pep	EWLLFGYRSDVWAKWLEMRQAGSPMTLLLAGTQIIDLKQSGVIPQDALQNDGDVFQTA					
	orf32a	EWLLFGYRSDVWAKWLEMRQAGSPLTLLLAGAXIIDLKQNGVIPQDALQNDGDVFQTA					
40		190	200	210	220	230	240
		250	260	270	280	290	300
45	orf32-1.pep	SVRLVKIPFVPQQDFDKLLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPOENVHLDKLH					
	orf32a	SVRLVKIPFVPQQDFDKLLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPOENVHLDKLH					
		250	260	270	280	290	300
		310	320	330	340	350	360
50	orf32-1.pep	AFWDKAHGFYTPETVSAHRRLSDDLNGGEALSATORLECQWTLQQHONGWRQGAEDWSRY					
	orf32a	AFWDKAHGFYTPETASAHRRLSDDLNGGEALSATORLECQWTLQQHONGWRQGAEDWSRY					
		310	320	330	340	350	360
55		370	380				
	orf32-1.pep	LFGQPSAPEKLAFAFVSKHQKIRX					
	orf32a	LFGQPSASEKLAFAFVSKHQKIRX					
		370	380				

60 Homology with a predicted ORF from *N.gonorrhoeae*

ORF32 shows 95.1% identity over a 82aa overlap with a predicted ORF (ORF32.ng) from *N. gonorrhoeae*:

orf32.pep	MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLP	57

orf32ng MVMNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLP 60
orf32.pep DVPCVHQDIHVRTWHSDAADIDTA 81
5 orf32ng DVPFVHQDIHVRTWHSDAADIDTAPVDAVIETFACDLPENVLNIIRRHKPLWLNWEYLS 120

An ORF32ng nucleotide sequence <SEQ ID 193> was predicted to encode a protein having amino acid sequence <SEQ ID 194>:

1 MVMNTYAFPV CWIFCKVIDN FGDIGVSWRL ARVLHRELGW QVHLWTDVVS
51 ALRALCPDLP DVPFVHQDIH VRTWHSDAAD IDTAPVDAV IETFACDLPE
101 NVLNIIRRHK PLWLNWEYLS AEESNERLHL MPSPQEGVQK YFWFMGFSEK
151 SGGLIRERDY REAVRFDTEA LRRRLVLPK NAPEWLLFGY RGDVWAKWLD
201 MWQQAGSLMT LLLAGAQIID SLKQSGVIPQ NALQNEGGVF QTASVRLVKI
251 PFVPQQDFDK LLHLADCAVI RGEDSFVRTQ LAGKPFWHI YPDENVHLD
301 KLHAFWDKAY GFYTPETASV HRLSDDLNG GEALSATORL ECGVL*

15 Further sequencing revealed the following DNA sequence <SEQ ID 195>:

1 ATGAATACAT ACGCTTTTCC TGTCTGTTGG ATTTTTTGCA AGGTCATCGA
51 CAATTTCCGGC GACATCGGCG TTTCGTGGCG GCTCGCCCGT GTTTTGCAACC
101 GCGAACTCGG TTGGCAGGTG CATTGTGGA CGGACGACGT GTCCGCCTTG
151 CGCGCGCTTT GTCCCGATTT GCCCGATGTT CCCTTCGTTC ATCAGGATAT
201 TCATGTCCGC ACTTGGCATT CCGATGCGGC AGACATTGAT ACCGCGCCCG
251 TTCCCGATGC CGTTATCGAA ACTTTTGCTT GCGACCTGCC CGAAAATGTG
301 CTGAACATCA TCCGCCGACA CAAACCGCTT TGGCTGAATT GCGAATATT
351 GAGCGCGGAG GAAAGCAATG AAAGGCTGCA CCTGATGCTT TCGCCGACAG
401 AGGGCGTTCA AAAATATTTT TGGTTTATGG GTTTCAGCGA AAAAAGCGGC
451 GGGTTGATAC GCGAACCGCA TTACCGCGAA GCCGTCCGTT TCGATACCGA
501 AGCCCTGCGC CGGCGGCTGG TGCTGCCCGA AAAAAACGCC CCCGAATGGC
551 TGCTTTTCGG CTATCGGGGC GATGTTTGGG CAAAGTGGCT GGACATGTGG
601 CAACAGGCAG GCAGCCTGAT GACCCTACTG CTGGCGGGGG CGCAAATTAT
651 CGACAGCCTC AAACAAAGCG GCGTTATTC GCAAAACGCC CTGCAAAATg
701 aaggcgGTGT CTTTCagacG gcatcgcTcC gccttGTCAA AatcCCGTTc
751 GTGCcGCAAC AGGAaCTCGA CAAATTGCTG CAcctcgcCG ACTGCGCCGT
801 GATACGCGGC GAAGACAGTT TCGTGCGTAC CCAGCTTGCC GGAACACCT
851 TTTTTTGGCA CATCTACCGC CAAGACGAGA ATGTCCATCT CGACAAACTC
901 CACGCCTTTT GGGATAAGGC ATACGGCTTC TACACGCCCG AAACCGCATC
951 GGTGCACCGC CTCCTTTCGG ACGACCTCAA CGGCGGAGAG GCTTTATCCG
1001 CAACACAACG CCTCGAATGT TGGCAAACCC TGCAACAACA TCAAAACGGC
1051 TGGCGGCAAG GCGCGGAGGA TTGAGCCGT TATCTTTTCG GGCAGCCTTC
1101 CGCATCCGAA AAACCTCGCCG CCTTTGTTTC AAAGCATCAA AAAATACGCT
1151 AG

40 This encodes a protein having amino acid sequence <SEQ ID 196; ORF32ng-1>:

1 MNTYAFPVCW IFCKVIDNFG DIGVSWRLAR VLHRELGWQV HLWTDVVSAL
51 RALCPDLPDV PFVHQDIHVR TWHSDAADID TAPVDAVIE TFACDLPENV
101 LNIIRRHKPL WLNWEYLSAE ESNERLHLM SPQEGVQKYF WFMGFSEKSG
151 GLIRERDYRE AVRFDTEALR RRLVLPKNA PEWLLFGYRG DVWAKWLDMW
201 QQAGSLMTLL LAGAQIIDSL KQSGVIPQNA LQNEGGVFQT ASVRLVKIPF
251 VPQQDFDKLL HLADCAVIRG EDSFVRTQLA GKPFWHIYP QDENVHLDKL
301 HAFWDKAYGF YTPETASVHR LLSDDLNGGE ALSATORLEC WQTLOQHONG
351 WRQGAEDWSR YLFGQPSASE KLAAFVSKHQ KIR*

ORF32ng-1 and ORF32-1 show 93.5% identity in 383 aa overlap:

50 orf32-1.pep 10 20 30 40 50 59
MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDV
orf32ng-1 MNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDVVSALRALCPDLPDV
10 20 30 40 50 60
55 orf32-1.pep 60 70 80 90 100 110 119
PCVHQDIHVRTWHSDAADIDTAPVDPVVIETFACDLPENVLHIIRRHKPLWLNWEYLSAE
orf32ng-1 PFVHQDIHVRTWHSDAADIDTAPVDAVIETFACDLPENVLNIIRRHKPLWLNWEYLSAE
60 70 80 90 100 110 120
120 130 140 150 160 170 179

	orf32-1.pep	ESNERLHLMPSPOEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRERLMLPEKNA
	orf32ng-1	ESNERLHLMPSPOEGVQKYFWFMGFSEKSGGLIRERDYREAVRFDTEALRRRLVLPEKNA
5		130 140 150 160 170 180
	orf32-1.pep	180 190 200 210 220 230 239
	orf32ng-1	SEWLLFGYRSDVWAKWLEMWRQAGSPMTLLLAGTQIIDSILKQSGVLPQDALQNDGDFVQT
10		190 200 210 220 230 240
	orf32-1.pep	240 250 260 270 280 290 299
	orf32ng-1	FEWLLFGYRGDVWAKWLDWQQAGSLMTLLLAGAQIIDSILKQSGVLPQNALQNEGGVFT
15		250 260 270 280 290 300
	orf32-1.pep	300 310 320 330 340 350 359
	orf32ng-1	HAFWDKAHGFYTPETVSAHRRLSDDLNGGEALSATQRLCEWQTLOQHONGWRQGAEDWSR
20		310 320 330 340 350 360
	orf32-1.pep	360 370 380
	orf32ng-1	YLFQPSASEKLAAFVSKHQKIRX
25		370 380

- 30 On this basis, including the RGD sequence in the gonococcal protein, characteristic of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF32-1 (42kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 35 7A shows the results of affinity purification of the His-fusion protein, and Figure 7B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA, giving a positive result. These experiments confirm that ORF32-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 24

- 40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 197>:

1 ..TTGTTCTGCTGTGNAAGTGGGCGTTTTTTCAGCAGTC CGGCGACGTG
51 GTTTCGGGNC AAAGACCCTG TAAATCAGGC GGTGTGCGG CTGTATNCGG
101 ACGAGTGGCG GCA.ACTTCG GTACGTTGGA AAATAGNCGC AACGTCGCAC
151 AGCCTGTGGC TCTGCACGCT GCTCGGAATG CTGGTGTCGG TATTGTTGCT
45 201 GCTTTTGGTG CGGCAATATA CGTTCAACTG GGAAAGCAGC CTGTTGAGCA
251 ATGCCGCTTC GGTACGCGCG GTGGAAATGT TGGCATGGCT GCCGTGCAAA
301 CTCGGTTTCC CTGTCCCGCA TCGCGGCTCG GTCATCGAAG GCCGTCTGAA
351 CGGCAATATT GCCGATGCGC GGGCTTGGTC GGGGCTGCTG GTCGNCAGTA
401 TCGCCTGCTA NGGCATCCTG CCGCGCCTG..

- 50 This corresponds to the amino acid sequence <SEQ ID 198; ORF33>:

1 ..LFLRVKVGFR FSSPATWFRX KDPVNQAVLR LYXDEWRXTS VRWKIXATSH
51 SLWLCTLLGM LVSLLLLLV RQYTFNWEST LLSNAASVRA VEMLAWLPSK
101 LGFPVPDARS VIEGRLNGNI ADARAWSGLL VXSIACXGIL PRL..

Further work revealed the complete nucleotide sequence <SEQ ID 199>:

```

      1 ATGTTGAATC CATCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGACGA
    51 AGGCGGTTTT ATTTTCAGCG GCGATCCCGT ACAGGCGACG GAGGCTTTGC
   101 GCCGCGTGGA CGGCAGTACG GAGGAAAAAA TCATCCGTCG GCGCGAGATG
  151 ATTGACAGGA ACCGTATGCT GCGGGAGACG TTGGAACGTG TGCCTGCGGG
   201 GTCGTTCTGG TTGTGGGTGG TGGCGGCGAC GTTTGCATT TTTACCGGTT
   251 TTTCAGTCAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTGT
   301 GTTTTGCGCG GCGTGTGGG CATGAATACG CTGATGCTGG CAGTATGGTT
   351 GGCAATGTTG TTCCTGCGTG TGAAAGTGGG GCGTTTTTTC AGCAGTCCGG
  401 CGACGTGGTT TCGGGGCAAA GACCCTGTAA ATCAGGCGGT GTTGGCGCTG
   451 TATGCGGACG AGTGGCGGCA ACCTTCGGTA CGTTGGA AAA TAGGCGCAAC
   501 GTCGCACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTCCGTAT
   551 TGTTGCTGCT TTTGGTGGG CAATATACGT TCAACTGGGA AAGCAGCTG
   601 TTGAGCAATG CCGCTTCGGT ACGCGCGGTG GAAATGTTGG CATGGCTGCC
  651 GTCGAAACTC GGTTCCTCTG TCCCGCATGC GCGGCGGTC ATCGAAGGCC
   701 GTCTGAACGG CAATATTGCC GATGCGCGGG CTTGGTCTGG GCTGCTGGTC
   751 GGCAGTATCG CCTGCTACGG CATCTGCCG CGCCTGCTGG CTTGGGTAGT
   801 GTGTAAATC CTTTTGAAA CAAGCGAAA CGGATTGGAT TTGAAAAGC
   851 CCTATTATCA GCGGCTCATC CGCGCTGGC AGAACAATA CACCGATGCG
  901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCACCGAAA TCATCTTGAA
   951 CGATGCGCCG AAATGGGCGG TCATGCTGGA GACCGAGTGG CAGGACGGCG
 1001 AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
 1051 ACCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
 1101 GGCGCAACTG CTTATCGGCG TGCGCGCCCA AACTGTGCCG GACCGCGGCG
 1151 TGTTGCGGCA GATTGTCCGA CTCTCGGAAG CGGCGCAGGG CGGCGCGGTG
 1201 GTGCAGCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
 1251 GGAACATTGG CGTAACGCGC TGGCCGAATG CGGCGCGGCG TGGCTTGAGC
 1301 CTGACAGGGC GGCGCAGGAA GGGCGTTTGA AAGACCAATA A

```

This corresponds to the amino acid sequence <SEQ ID 200; ORF33-1>:

```

    30      1 MLNPSRKLVE LVRILDEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAEM
      51 IDNRNMLRET LERVAGSEFW LWVVAATFAF FTGFSVTYLL MDNQGLNFFL
   101 VLAGVLGMNT LMLAVWLAML FLRVKVGRRFF SSPATWFRGK DPNVQAVLRL
   151 YADEWRQPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
   201 LSNAASVRVAV EMLAWLPSKL GFPVPDARAV IEGRNLGNIA DARAWSGLLV
   251 GSIACYGILP RLLAWVVCKI LLKTSENGLD LEKPYQYAVI RRWNKITDA
   301 DTRRETVS AV SPKILNDAP KWAVMLETEW QDGEWFEGRL AQEWLDKGVA
   351 TNREQVALE TELKQKPAQL LIGVRAQTPV DRGVLRLQIVR LSEAAQGGAV
   401 VQLLAEQGLS DDLSEKLEHW RNALAECEGAA WLEPDRAAQE GRLKDQ*

```

Computer analysis of this amino acid sequence gave the following results:

40 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF33 shows 90.9% identity over a 143aa overlap with an ORF (ORF33a) from strain A of *N. meningitidis*:

```

      45      orf33.pep                                10      20      30
      orf33a                                LFLRVKVGRRFFSSPATWFRXKDPVNQAVLR
      90      100      110      120      130      140
      50      orf33.pep                                40      50      60      70      80      90
      orf33a                                LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVRQYTFNWESTLLSNAASVRA
      150      160      170      180      190      200
      55      orf33.pep                                100      110      120      130      140
      orf33a                                VEMLAWLPSKLGFPVPDARSVIEGRNLGNIA DARAWSGLLVXSIACXGILPRL
      210      220      230      240      250      260
      60      orf33a                                VEMLAWLPAKLGFPVPDARAVIEGRNLGNIA DARAWSGLLVGSIACYGILPRLAWAVCK
      270      280      290      300      310      320
      orf33a                                ILXXTSENGLDLEKXXXXXIRRWNKITDADTRRETVS AVSPKIVLNDAPKWAVMLETE

```

The complete length ORF33a nucleotide sequence <SEQ ID 201> is:

```

1 ATGTTGAATC CATCCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGAAGA
51 AGGCGGCTTT ATTTTCAGCG GCGATCCCGT GCAGGCGACG GAGGCTTTGC
101 GCCGCGTGGA CGGCAGTACG GAGGAAAAAA TCATCCGTCG GCGGAAGATG
5 151 ATCGACAGGA ACCGTATGCT GCGGGAGACG TTGGAACGTG TCGGTGCGGG
201 GTCGTTCTGG TTGTGGGTGG CGGCGGCGAC GTTTGCGTTT NTTACCGNTT
251 TTTCAGTTAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTGG
301 GTTTTGGCGG GCGTGTGTTG CATGAATACG CTGATGCTGG CAGTATGGTT
351 GGCAATGTTG TTCCTGCGCG TGAAAGTGGG GCGTTTTTTC AGCAGTCCGG
10 401 CGACGTGTTT TCGGGGCAAA GACCCTGTCA ATCAGGCGGT GTTGGCGCTG
451 TATGCGGACG AGTGGCGGCN ACCTTCGGTA CGTTGGAAAA TAGGCGCAAC
501 GTCGCACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTGCGTAT
551 TGTTGCTGCT TTTGGTGGCG CAATATACGT TCAACTGGGA AAGCAGCTG
601 TTGGGCGATT CGTCTTCGGT ACGGCTGGTG GAAATGTTGG CATGGCTGCC
15 651 TGCGAACTG GGTTCCTCCG TGCCTGATGC GCGGGCGGTC ATCGAAGGTC
701 GTCGAACTG CAATATTGCC GATGCGCGGG CTTGGTGGG GCTGCTGGTC
751 GGCAGTATCG CCTGCTACGG CATCCTGCGG CGCCTCTTGG CTTGGGCGGT
801 ATGCAAAATC CTNTTGNAAA CAAGCGAAAA CGGCTTGAT TTGAAAAAGC
851 NCNNNNNTCN NNCGNTCATC CGCCGCTGGC AGAACAAAA CACCGATGCG
20 901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCGCCGAAAA TCGTCTTGAA
951 CGATGCGCGG AAATGGGCGG TCATGCTGGA GACCGAATGG CAGGACGGCG
1001 AATGGTTCGA GGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
1051 GCCAATCGGG AACAGTTGCG CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
1101 GCGCAACTG CTTATCGGCG TGCGCGCCCA AACTGTGCCC GACCGCGGCG
25 1151 TGTTGCGGCA GATCGTCCGA CTTTCGGAAG CGGCGCAGGG CGGCGCGGTG
1201 GTGCANCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAGCT
1251 GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTGGAAC
1301 CCGACAGAGC GGCAGGAA GGCCGTCTGA AAACCAACGA CCGCACTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 202>:

```

30 1 MLNPSRKLVE LVRILEEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAKM
51 IDNRNMLRET LERVAGSEF LWVAAATFAF XTXFSVTYLL MDNQGLNFFL
101 VLAGVXGMNT LMLAVWLAML FLRVKVGRRF SSPATWFRGK DPNVQAVLRL
151 YADEWRXPSV RWKIGATSHS LWLCTLLGML VSVLLLLLV RQYTFNWESTL
201 LGDSSSVRLV EMLAWLPAKL GFPVPDARAV IEGRNLGNIA DARAWSGLLV
35 251 GSIACYGILP RLLAWAVCKI LXXTSENGLD LEKXXXXXXI RRWQNKITDA
301 DTRRETSAV SPKIVLNDAP KWAVMLETEW ODGEWFEGRL AQEWLDKGVA
351 ANREQVALE TELKOKPAQL LIGVRAQTVP DRGVLRQIVR LSEAAQGGAV
401 VXLLEQGLS DDLSEKLEHW RNALTECGAA WLEPDRAAQE GRLLTNDRT*

```

ORF33a and ORF33-1 show 94.1% identity in 444 aa overlap:

```

40 10 20 30 40 50 60
orf33a.pep MLNPSRKLVELVRILEEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAKMIDNRNMLRET
orf33-1 MLNPSRKLVELVRILDEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAEMIDNRNMLRET
10 20 30 40 50 60
45 70 80 90 100 110 120
orf33a.pep LERVAGSEFWLWVAAATFAFXTXFSVTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML
orf33-1 LERVAGSEFWLWVAAATFAFFTGFSTYLLMDNQGLNFFLVLAGVLMNTLMLAVWLAML
70 80 90 100 110 120
50 130 140 150 160 170 180
orf33a.pep FLRVKVGRRFSSPATWFRGKDPVNQAVLRLYADEWRXPSVRWKIGATSHSLWLCTLLGML
orf33-1 FLRVKVGRRFSSPATWFRGKDPVNQAVLRLYADEWRQPSVRWKIGATSHSLWLCTLLGML
130 140 150 160 170 180
55 190 200 210 220 230 240
orf33a.pep VSVLLLLLVRYTFNWESTLLGDSSSVRLVEMLAWLPAKLGFVPDARAVIEGRNLGNIA
orf33-1 VSVLLLLLVRYTFNWESTLLSNAASVRAVEMLAWLPSKLGFPVPDARAVIEGRNLGNIA
190 200 210 220 230 240
60 250 260 270 280 290 300
orf33a.pep DARAWSGLLVGSIACYGILPRLLAWAVCKILXXTSENGLDLEKXXXXXXIIRWQNKITDA

```

-158-

	orf33-1	DARAWSGLLVGSIA	250	260	270	280	290	300
			310	320	330	340	350	360
5	orf33a.pep	DTRRET	310	320	330	340	350	360
	orf33-1	DTRRET	310	320	330	340	350	360
10	orf33a.pep	TELKQK	370	380	390	400	410	420
	orf33-1	TELKQK	370	380	390	400	410	420
15	orf33a.pep	RNALTE	430	440	450			
	orf33-1	RNALTE	430	440				

Homology with a predicted ORF from *N.gonorrhoeae*

ORF33 shows 91.6% identity over a 143aa overlap with a predicted ORF (ORF33.ng) from *N.gonorrhoeae*:

25	orf33.pep	LFLRVKVG	30
	orf33ng	LMDNQGLN	100
30	orf33.pep	LYXDEWR	90
	orf33ng	LYADQWR	160
	orf33.pep	VEMLAWL	143
35	orf33ng	VEMLAWL	220

An ORF33ng nucleotide sequence <SEQ ID 203> was predicted to encode a protein having amino acid sequence <SEQ ID 204>:

	1	MIDRDRMLRD	TLERV	RAGSF	WLWVV	VASMM	FTAGF	SGTYL	LMDNQ	GLNFF
	51	LVL	AGVL	GMN	TLML	AVLAT	LFLRV	KVGRF	FSS	PATWFRG
40	101	LYADQ	WRQPS	VRWK	IGATAH	SLWL	CTLL	GM	LVS	VLLLLLV
	151	LLS	NAASVRA	VEML	AWLPSK	LGFP	VPDARA	VIEGR	LNGNI	ADARAW
	201	VGS	IVCYGIL	PRLL	AWVVK	ILLK	TSENGL	DLEK	TYQAV	IRRWQNKITD
	251	ADT	RRET	VSA	VSPK	IVLND	PKW	LMLE	TE	WQDQGWFEGR
	301	AAN	REQVAAL	ETEL	KQKPAQ	LLIG	VRAQTV	PDRG	VLRQIV	RLSEAAQGG
45	351	VVQ	LLAEQGL	SDDL	SEKLEH	WRNAL	TECGA	AWLE	PDRVAQ	EGRLKDQ*
	1	ATGTT	GaatC	CATCC	GaaAA	ACTG	gttgag	ctGg	TCCgtA	Ttttgaataa
	51	agggg	gtTTT	atttt	cagcg	gcat	cctgt	gcagg	cgacg	gaggctttgc
	101	gccc	gcgtgga	cggc	AGTACG	GAgg	Aaaaaa	tctt	ccgtcg	GGCGGAGAtg
50	151	atcg	ACAGGg	accgt	atggt	gcggg	ACaCg	Ttgg	aacGTG	TGCGTGCGgg
	201	gtcgt	TctgG	TTAT	GGGTGG	Tggt	ggCAtC	gATG	ATGTt	acCCGCCGAT
	251	TTT	CAGgcac	ttat	CttCTG	ATGG	ACaatC	AGGG	GcTGAA	TtTCTTTT
	301	GTTT	TggcgG	GAGT	GtggG	CATG	aatacG	ctgA	TGCTGG	CAGTATGgtt
	351	gGCA	ACGTTG	TTC	TGCGCG	TGAA	AGTGGG	ACGG	TTTTTC	AGCAGTCCGG
55	401	CGAC	TGGT	TTCGGG	GCAAA	GGCC	CTGTAA	ATC	AGGCGGT	GTTGCGGCTG
	451	TATG	CGGACC	AGT	GCGGCA	ACCT	TCGGTA	CGAT	GGA	AAATAGGCGCAAC
	501	GGCG	CACAGC	TTGT	GGCTCT	GCAC	GCTGCT	CGGA	ATGCTG	GTGTCGGTAT
	551	TGCT	GCTGCT	TTT	GGTGGG	CAAT	ATACGT	TCA	ACTGGGA	AAGCACGCTG
	601	TTG	AGCAATG	CCG	CTCGGT	ACGC	GCGGTG	GAA	ATGTTGG	CATGGCTGCC
60	651	GTC	GAACTC	GGT	TTCCCTG	TCCC	GATGC	GCGG	GCGGTC	ATCGAAGGTC
	701	GTCT	GAACGG	CAAT	ATTGCC	GATG	CGCGGG	CTTG	GTGCGG	GCTGCTGGTC
	751	GGC	AGTATCG	TCTG	TACGG	CATC	CTGCGG	CGCT	CTTGG	CTTGGGTAGT

Further sequence analysis revealed the following DNA sequence <SEQ ID 205>:

801 GTGTAATAATC CTTTGTAAAA CAAGCGAAAA CGGattgGAT TTGGAAAAAA
 851 CCTATTATCA GGCGGTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
 901 GATACGCGTC GGGAAACCGT GTCCGCCGTT TCGCggaAAA TCGTCTTGAA
 951 CGATGCGCCG AAATGGGCGC TCATGCTGGA GACCGAGTGG CAGGACGGCC
 1001 AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
 1051 GCCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
 1101 GGCGCAACTG CTTATCGGCG TACGCGCCCA AACTGTGCCG GACCGGGGCG
 1151 TGCTGCGGCA GATTGTGCGG CTTTCGGAAG CGGCGCAGGG CGGCGCGGTG
 1201 GTGCAGCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
 1251 GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTTGAGC
 1301 CTGACAGGGT GGCGCAGGAA GGCCGTTTGA AAGACCAATA A

This encodes a protein having amino acid sequence <SEQ ID 206; ORF33ng-1>:

1 MLNPSRKLVE LVRILNKGGF IFSGDPVQAT EALRRVDGST EEKIFRRAEM
 51 IDRDRMLRDT LERVAGSEFW LWVVVASMMF TAGFSGTYLL MDNQGLNFFL
 101 VLAGVLGMNT LMLAVWLATL FLRVKVGRRF SSPATWFRGK GPVNQAVLRL
 151 YADQWRQPSV RWKIGATAHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
 201 LSNAASVRV EMLAWLPSKL GFPVPDARV IEGRNLGNIA DARAWSGLLV
 251 GSIVCYGILP RLLAWVVKI LLKTSENGLD LEKTYQAVI RRWQNKITDA
 301 DTRRETSAV SPKIVLNDAP KWALMLETEW QDGQWFEGR L AQEWLDKGVA
 351 ANREQVALE TELKQKPAQL LIGVRAQTVP DRGVLRQIVR LSEAAQGGAV
 401 VQLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRVAQE GRLKDQ*

ORF33ng-1 and ORF33-1 show 94.6% identity in 446 aa overlap:

25	orf33-1.pep	10	20	30	40	50	60
	orf33ng-1	10	20	30	40	50	60
30	orf33-1.pep	70	80	90	100	110	120
	orf33ng-1	70	80	90	100	110	120
35	orf33-1.pep	130	140	150	160	170	180
	orf33ng-1	130	140	150	160	170	180
40	orf33-1.pep	190	200	210	220	230	240
	orf33ng-1	190	200	210	220	230	240
45	orf33-1.pep	250	260	270	280	290	300
	orf33ng-1	250	260	270	280	290	300
50	orf33-1.pep	310	320	330	340	350	360
	orf33ng-1	310	320	330	340	350	360
55	orf33-1.pep	370	380	390	400	410	420
	orf33ng-1	370	380	390	400	410	420
60	orf33-1.pep	430	440				
	orf33ng-1	430	440				

orf33ng-1

RNALTECGAAWLEPDRVAQEGRLKDQX
430 440

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is
5 predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be
useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 25

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 207>:

```

10      1  ..CAGAAGAGTT TGTCGAGAAT TTCTTTATGG GGTTTGGGCG GCGTGTTTTT
      51  CGGGGTGTCC GGTCTGGTAT GGTTTTCTTT GGGCGTTTCT TT.GAGTGCG
     101  CCTGTTTTC GGGTGTTCCT TTTCGGGGTT CGGGACGGGG GACGTTTGTG
     151  GGCAGTACGG GGGTTTCTTT GAGTGTGTTT TCAGCTTGTG TTCC.GGCCT
     201  CGTCCGGCTG CCTGTCGGTT TGAGCTGTGT CGGCAGGTTG CG..GTTTGA
     251  CCCGGTTTTT CTTGGGTGCG GCAGGGGACG TCATTCTCCT GCCGCTTTCG
     15  301  TCTGTGCCGT CCGGCTGTGC GGGTTCGGAT GAGGCGGCGT GGTGGTGTTG
     351  GGGTTGGGCG GCATCTTGTT CCGACTACGC CGTTTGGCAG CCAGAATTCG
     401  GTTTCGCGGG GGCTGTCGGT GTGTTGCGGT TCGGCTTGAA GGGTTTTGTC
     451  GTCC..

```

This corresponds to the amino acid sequence <SEQ ID 208; ORF34>:

```

20      1  ..QKSLSRISLW GLGGVFFGVVS GLVWFSLGVS XECACFSGVVS FRGSGRGTFV
     51  GSTGVSLSVF SACVXGVVRL PVGLSCVGR LXXLTRFFLGA AGDVILLPLS
     101  SVPSGCAGSD EAAWCSGWA ASCPTTFFGS QNSVSRGLSV CCGSA*RVLS
     151  S..

```

Further work revealed the complete nucleotide sequence <SEQ ID 209>:

```

25      1  ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCKGGTG TGCCTGCCGT
     51  GCCGGGTCAG AATAGGTTGT CCAGAATTTT TTTATGGGGT TTGGGCGGCG
     101  TGTTTTTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTG
     151  GGCTGCGCCT GTTTTTCGGG TGTTCCTTTT CGGGGTTTCG GACGGGGGAC
     201  GTTTGTGGGC AGTACGGGGG TTTCTTTGAG TGTGTTTCA GCTTGTGTTC
     30  251  CGGCGTCGTC CGGCTGCCTG TCGGTTGAG CTGTGTCGGC AGGTTGCGGT
     301  TTGACCCGGT TTTCTTTGGG TCGGCGAGGG GACGGCAGTC CGCTGCCGCT
     351  TTCGTCTGTG CCGTCCGGCT GTGCGGGTTC GGATGAGGCG GCGTGTGGT
     401  GTTCGGGTTG GCGGCACTCT TGTCCGACTA CGCCGTTTGG CAGCCAGAAT
     451  TCGGTTTCGC GGGGGCTGTC GGTGTGTTGC GGTTCGGCTT GAAGGGTTTT
     35  501  GTCGCCGTTT GGGTTGAATG TGCTGACGAT GCCTATTGCC AATGCGCCGA
     551  TGGCGGCGAT ACAGATGAGC AATACGGCGC GTATCAGGAG TTTGGGGGTC
     601  AGCCTGAAGG GTTTGTTTCG TTTTTCGCTT ATTTTGATTG TGCTTTTGGG
     651  GTGTGCGGCA ATGCCGCTCT AAGGCGGTTT AGACGGCATT GCCGAGTCAG
     701  CGTTGGACGT AGTTTGGGTA GAGGGTGATG ACTTTTGTGA CGCCGACGGT
     40  751  GGTGCTGACT TTTTGGGTAA TCTGCGCCTG TTCTTCGGGG GTGAGGATGC
     801  CCATAACGTA GGTACGTTG CCGTAGGTAA CGATTTTGAC GCGCGCCTGT
     851  GTGGCGGGGC TGATGCCCAA CAGCGTGCGC CGGACTTTGG ATGTGTTCCA
     901  AGTGTGCGCG GCGATGTCGC CGGCAGTGCG CGGCAGGGAG GCGACGGTAA
     951  TATAGTTGTA CACGCCCTCG GCGGCCTGTT CGGAACGTGC AATCTGACCG
     45  1001  ACGAACTGTT TTTCCGCTTC GGTGGCGACT TGTCGAGCA GCAGCAGGTG
     1051  GCGGTTGTAG CCGACGACGG AGATTTGGGG CGTGTAGCCT TTGGTTTGGT
     1101  TGTTTGGGCG CAGATAGGAA CGGGCGGTGG TTTGATACG CAACGCCATA
     1151  ACGTTGTCGT CGGTTTGC GC CGGCGGTGTT CGGCGGTGCA CGGCGGATTT
     1201  CGGCGCGACG GCGGCGCTTC CGATTACTGC GCTGACGACG CCCTAAGGG
     50  1251  CAAGGCTGAA AATGGCGGCA ATCAGGGTGC GGACGGTGTG CGGTTTGGGT
     1301  TTCATCGGGT GCTTCCTTTC TTGGGCGTTT CAGACGGCAT TGCTTTCGCG
     1351  CATGCCGTCT GA

```

This corresponds to the amino acid sequence <SEQ ID 210; ORF34-1>:

```

55      1  MMMPFIMLPW IAGVPAVPGQ NRLSRISLWG LGGVFFGVSG LVWFSLGVSL
     51  GCACFSGVSF RSGRGTFVG STGVSLSVFS ACVPASSGCL SV*AVSAGCG
     101  LTRFFLGAAG DGSPLPLSSV PSGCAGSDEA AWWCSGWAAS CPTTFFGSQN
     151  SVSRGLSVCC GSA*RVLSPF GLNVLTMPIA NAPMAAIQMS NTARIRSLGV

```

201 SLKGLFGFFA ILIVLLGCRA MPSEGGSDGI AESALDVVLV EGDDFLYADG
 251 GADFLGNLRL FFGGEDAHNV GYVAVGNDFD ARLCGGADAQ QRGADFGCVP
 301 SVAGDVAGSA RQGGDGNIVV HAFGGLFGTC NLTDLEFFAF GGDLSEQQQV
 351 AVVADDGDLG RVAFLVLVLA QIGTGGGFDI QRHNVVVGLR AGGSAVDGGF
 401 RADGGASDYC ADAAAKGKAE NGGNQAGDGV RFGFHRVLPF LGVSDGIALR
 451 HAV*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF34 shows 73.3% identity over a 161aa overlap with an ORF (ORF34a) from strain A of *N.*

10 *meningitidis*:

		10	20	30	
orf34.pep		QKSLSRISLWGLGGVFPVSGVLVWFSLGVSXE-----CAC			
orf34a	MMXPXIMLPWIAGVPAVPGQRLSRXSLWGLGGXFFGVSGVLVWFSLGVSXSXGVCAC	10	20	30	40
15		40	50	60	70
orf34.pep	FSGVSFRGSGRGTFFVGSTGVSLSVFSACVXGVVRLPVGLSCVGRLLX-----LTRFFLGA				
20	orf34a	FSGVSFRGSGRGTFFVGSTGVSLSVFSACA-----PASSGCLSVXAVSAGCGLTRFXGA	70	80	90
		100	110	120	130
25	orf34.pep	AGDVILLPLSSVPSGCAGSDEAAWCSGWAASCPTTTPFGSQNSVSRGLSVCCGSAXRVLS			
	orf34a	AGDGSPLPLSSVPSGCAGADEEAXXCSGWAASCPTTTPFGSQNSVSRGLSVCCGSVWRVLS	120	130	140
		150	160	170	180
30	orf34.pep	S			
	orf34a	PFGXNVLTMPIANAPMAVIQMSNTARIRSLGVSLKGLFXFFAILIVLLGCRAMPSEGGSD	180	190	200
		210	220	230	

The complete length ORF34a nucleotide sequence <SEQ ID 211> is:

35	1	ATGATGATNC	CGTTNATAAT	GCTTCCTTGG	ATTGCGGGTG	TGCCTGCCGT
	51	GCCGGGTCAG	AAGAGGTTGT	CGAGAANTTC	TTTATGGGGT	TTAGGCGGCN
	101	TGTTTTTCGG	GGTGTCCGGT	TTGGTATGGT	TTTCTTTGGG	CGTTTCTNNT
	151	TCTTTGGGTG	TTTCTNTGGG	CTGTGCCTGT	TTTTCGGGTG	TTTCTTTTCG
	201	GGGTTTCGGG	CGGGGGACGT	TTGTGGGCAG	TACNGGGGTT	TCTTTGAGTG
40	251	TGTTTTTCAG	TTGTGCTCCG	GCGTCGTCGG	GCTGCCTGTC	GGTTTNAAGT
	301	GTGTCCGCAG	GTTGCGGTTT	GACCCGGNTT	TTCTTNGGTG	CGGCAGGGGA
	351	CGGCAGTCCG	CTGCCGCTTT	CGTCTGTGCC	GTCCGGCTGT	GCGGGTGCGG
	401	ATGAGGAGGC	GTNGTNGTGT	TCGGGTTGGG	CGGCATCTTG	TCCGACTACG
	451	CCGTTTGGCA	GCCAGAATTC	GGTTTCGCGG	GGGCTGTGCG	TGTGTGCGGG
45	501	TTCGGTNTGG	AGGGTTTGTG	CNCCGTTCCG	GTNGAATGTG	CTGACGATGC
	551	CTATTGCCAA	TGCGCCGATG	GCGGTGATAC	AGATGAGCAA	TACGGCGCGT
	601	ATCAGGAGTT	TGGGGGTCAG	CCTGAAGGGT	TTGTTCNGTT	TTTTTGCCAT
	651	TTTGATTGTG	CTTTTGGGGT	GTCCGGCAAT	GCCGTCTGAA	GGCGGTTTCAG
	701	ACGGCATTCG	CGAGTCAGCG	TTGGACGTAG	TTTNGGTAGA	GGGTGATGAC
50	751	TTTTTGTACG	CCGACGGTGG	TGCTGACTTT	TTGGGTAATC	TGCGCCTGTT
	801	CTTCGGGGGT	GAGGATGCCC	ATAACGTAGG	TTACGTTGCC	GTAGGTAACG
	851	ATTTTGACGC	GCGCCTGTGT	GGCGGGGCTG	ATGCCCAACA	GCGTGGCGCG
	901	GACTTTGGAT	GTGTTCCAAG	TGTCGCGGCG	GATGTCGCGG	GCAATGCGCG
	951	GCAGGGAGGC	GACGGTAATG	TANTTGTACA	CGCCTTCGGC	GGCCTGTTTC
55	1001	GAACGTGCAA	TCTGACCGAC	GAACGTGTTT	TCGCCTTCGG	TGGCGACTTG
	1051	TCCGAGCAGC	AGCAGGTGGC	GGTTGTAGCC	GACAACGGAG	ATTTGGGGCG
	1101	TGTANCTTTT	GGTTTGGTTG	TTTTGGCGCA	GATAGGAGCG	GGCGGTGGTT
	1151	TCGATACGCA	GCGCCATTAC	GTTGTCGTCT	GTTNGCGCGC	CGGTGGTTTC
	1201	GCGGTGACAG	GCGGATTTTC	CGCCGACCGC	CGCGCCGCGG	ACGACTGCGC
60	1251	TGACGCAGCC	GCCGAGGGCA	AGGCTGAGGA	CGGCGGCAGT	CAGGGTGCGG
	1301	ACGGTGTGCG	GTTTGGGTTT	CATCGGGTGC	TTCTTTTCTT	GGGCGTTTCA
	1351	GACGGCATTG	CTTTGCGCCA	TGCCGTCTGA		

ORF34a and ORF34-1 show 91.3% identity in 459 aa overlap:

Homology with a predicted ORF from *N.gonorrhoeae*

orf34.pep QKSLSRISLWGLGGVFFGVSGLVWFSLGVSXE-----CAC 35

	orf34ng	MMMPFIMLPWIAGVPAVPGQKRLSRISLWGLAGVFFGVSGLVWFSLVGSFSLGVSLGCAC	60
	orf34.pep	FSGVSFRGSGRGT FVGSTGVSLSVFSACVXGVVRLPVGLSCV-----GRLXXLTRFFLGA	90
5	orf34ng	FSGVSFRGSGWGA FVGSTGVSLSVFSACVP----VPVNESAARAASEGR--GLTRFFLGA	114
	orf34.pep	AGDVILLPLSSVPSGCAGSDEAAWWSGWAASCTTPFGSQNSVSRGLSVCCGSAXRVLS	150
10	orf34ng	AGDGSPLPLSSVPSGCAGSDEAAWWSGWAASCTAPFGSQNSVSRGLSVCCGSVWRVLS	174
	orf34.pep	S	175
	orf34ng	PFGNLVLTMTANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSD	234

15 The complete length ORF34ng nucleotide sequence <SEQ ID 213> is:

	1	ATGATGATGC	CGTTCATAAT	GCTTCCTTGG	ATTGCGGGTG	TGCCTGCCGT
	51	GCCGGGTC	AAGAGGTTGT	CGAGAATCTC	TTTATGGGT	TTGGCCGGCG
	101	TGTTTTTC	GGTGTCGGT	TTGGTATGGT	TTTCTTTGGG	CGTTTCTTTT
	151	TCTTTGGGT	TTTCTTTGGG	CTGCGCCTGT	TTTTCGGGTG	TTTCTTTTCG
20	201	GGGTTCCGGA	TGGGGGGCGT	TTGTGGGCAG	TACGGGGGTT	TCTTTGAGTG
	251	TGTTTTTCAGC	TTGTGTTCCG	GTGCCGGTTA	ACGAATCGGC	TGCCCGGGCC
	301	GCATCCGAAG	GGCGCGGTTT	gACCCGGTTT	TTCTTGGGTG	CGGCAGGGGA
	351	CGGCAGTCCG	CTGCCGCTTT	CTTCTGTGCC	GTCCGGCTGT	GCGGGTTCGG
	401	ATGAGGCGGC	GTGGTGGTGT	TCGGGTGGG	CGGCATCTTG	TCCGACGGCG
25	451	CCGTTTGGA	CCAGAATTC	GGTTTCGCGG	GGGCTGTCGG	TGTGTTGCGG
	501	TTCGGTTTGG	AGGGTTTTGT	CGCCGTTCGG	GTGAATGTG	CTGACGATGC
	551	CTACTGCCAA	TGCCCGCATG	GCGGTGATAC	AGATGAGCAA	TACGGCGCGT
	601	ATCAGGAGTT	TGGGGGTGAG	CCTGAAGGT	TTGTTGCGTT	TTTTTGCCAT
	651	TTTGATTGTG	CTTTTGGGGT	GTCCGGCAAT	GCCGTCTGAA	GGCGGTTTAC
30	701	ACGGCATTGC	CGAGTCAGCG	TTGGACGTAG	TTTTGGTAGA	GGGTAATGAC
	751	TTTTTGATAC	CCGACggtGG	TGCTGACTTT	TTGGGTAATC	TGCGCCTGTT
	801	CTTCGGGGGT	GAGGATGCCC	ATAACGTAGG	TTACATTGCC	GTAGGTAATG
	851	ATTTTGACGC	GCGCCTGTGT	AGCGGGGCTG	ATGCCAGCA	GcgtgCGCGC
	901	GACTTTGGAC	GTGTTCCAAG	TGTCGCCGGC	GATGTCGCCC	GCACTGCGCG
35	951	GCAGGGAGGC	GACGGTAATG	TAGTTGTATA	CGCCTTCGGC	GGCCTGTTTC
	1001	GAACGTGCAA	TCTGACCGAC	GAAGTGTGTT	TGCGCTTCGG	TGGCGACTTG
	1051	TCCGAGCAGC	AGCAGGTGGC	GGTTGTAGCC	GACGACGGAG	ATTTGGGGCG
	1101	TGTAGCCTTT	GGTTTGGTTG	TTTTGGCGCA	GGTAGGAACG	GGCGGTGGTT
	1151	TCGATACGCA	ACGCCATAAC	GTtgtCATCG	GTTtgcgcg	CGGTGGTTcg
40	1201	gCGGTCGATG	ACGGATTTTG	CGCCGACGGC	GGCCCCGCGC	ACGACTGCGC
	1251	TGAAGCAGCC	GCCGAGGGCA	AGGCTGAGGA	CGGCGCAAT	CAGGGTGC
	1301	ACGGTGTGTG	GTTTGGGTTT	CATCGGGGAC	TTCTTTCTT	GGGCGTTTCA
	1351	GACGCGATTG	CTTTGCGCCA	TGCCGTCTGA		

This encodes a protein having amino acid sequence <SEQ ID 214>:

45	1	MMMPFIMLPW	IAGVPAVPGQ	KRLSRISLWG	LAGVFFGVSG	LVWFSLVGSF
	51	SLGVSLGCAC	FSGVSFRGSG	WGA FVGSTGV	SLSVFSACVP	VPVNESAARA
	101	ASEGRGLTRF	FLGAAGDGSP	LPLSSVPSGC	AGSDEAAWWS	SGWAASCTPA
	151	PFGSQNSVSR	GLSVCCGSVW	RVLSPFGLNV	LTMPTANAPM	AVIQMSNTAR
	201	IRSLGVSLKG	LFGFFAILIV	LLGCRAMPSE	GGSDGIAESA	LDVVIVEGND
50	251	FLYADGGADF	LGNLRLFFGG	EDAHNVGYIA	VGNDFDARLC	SGADAQQRGA
	301	DFGRVPSVAG	DVARSAQQG	DGNVVVYAFG	GLFGTCNLTD	ELFFAFGGDL
	351	SEQQQVAVVA	DDGDLGRVAF	GLVVLQVGT	GGGFDQQRHN	VVIGLRAGGS
	401	AVDDGFCADG	GPADDCAEAA	AEGKAEDGGN	QGADGVWFGF	HRGLPFLGVS
	451	DGIALRHAV*				

55 ORF34ng and ORF34-1 show 90.0% identity in 459 aa overlap:

		10	20	30	40	4	50
	orf34-1.pep	MMMPFIMLPWIAGVPAVPGQNRSLRISLWGLGGVFFGVSGLVWFSLVGS-----LGACAC					
	orf34ng	MMMPFIMLPWIAGVPAVPGQKRLSRISLWGLAGVFFGVSGLVWFSLVGSFSLGVSLGCAC					
60		10	20	30	40	50	60
	orf34-1.pep	FSGVSFRGSGRGT FVGSTGVSLSVFSACVPASSGCLSVXAVSAGCGLTRFFLGAAGDGSP					
65	orf34ng	FSGVSFRGSGWGA FVGSTGVSLSVFSACVPVPVNESAARAASEGRGLTRFFLGAAGDGSP					

-164-

		70	80	90	100	110	120
		120	130	140	150	160	170
5	orf34-1.pep	LPLSSVPSGCAGSDEAAWCSGWAASCPTTFFGSQNSVSRGLSVCCGSAXRVLSPFGLNV					
	orf34ng	LPLSSVPSGCAGSDEAAWCSGWAASCPTTFFGSQNSVSRGLSVCCGSVWRVLSPFGLNV					
		130	140	150	160	170	180
10	orf34-1.pep	LTMPIANAPMAAIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA					
	orf34ng	LTMTANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA					
		190	200	210	220	230	240
15	orf34-1.pep	LDVVLVEGDDFLYADGGADFLGNLRLFFGGEDAHNVGYVAVGNDFDARLCGGADAQQRGA					
	orf34ng	LDVVLVEGDDFLYADGGADFLGNLRLFFGGEDAHNVGYIAVGNDFDARLCGGADAQQRGA					
		250	260	270	280	290	300
20	orf34-1.pep	DFGCVPSVAGDVAGSARQGGDGNIVVHAFGGFLGTCNLTDDELFFAFGGDLSEQQVAVVA					
	orf34ng	DFGRVPSVAGDVARSARQGGDGNVVVYAFGGFLGTCNLTDDELFFAFGGDLSEQQVAVVA					
25		310	320	330	340	350	360
	orf34-1.pep	DDGDLGRVAFGLVVLAAQIGTGGGFDTORHNVVGLRAGGSADVDDGFRADGGASDYCADAA					
	orf34ng	DDGDLGRVAFGLVVLAAQVGTGGGFDTORHNVVIGLRAGGSADVDDGFCADGGPADDCAEAA					
30		370	380	390	400	410	420
	orf34-1.pep	AKGKAENGGNQGADGVRFGRVLPFLGVSDGIALRHAVX					
	orf34ng	AEGKAEDGGNQGADGVWFGFHRGLPFLGVSDGIALRHAVX					
35		430	440	450	460		

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 26

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 215>:

```

1  ATGAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT
51 CGCCGCCTGC GGATT.CAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
101 CCGCCGCCGA CAACGGCGCG GCGTAAAAAA GAAATCGTCT TCGGCACGAC
151 CGTCGGCGAC TTCGGCGATA TGGTCAAAGA ACAATCCAA GCCGAGCTGG
201 AGAAAAAAGG CTACACCGTC AACTGGTCG AGTTACCGA CTATGTACGC
251 CCGAATCTGG CATTGGCTGA GGGCGAGTTG

```

This corresponds to the amino acid sequence <SEQ ID 216; ORF4>:

```

1  MKTFFKTLA AALALILAAC G.QKDSAPAA SASAAADNGA AKKEIVFGTT
51 VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GEL

```

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 217>:

```

1  ATGAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT
55 51 CGCCGCCTGC GCGGCTCAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
101 CCGCCGCCGA CAACGGCGCG GCGAAAAAAG AAATCGTCTT CGGCACGACC
151 GTCGGCGACT TCGGCGATAT GGTCAAAGAA CAAATCCAAG CCGAGCTGGA
201 GAAAAAAGGC TACACCGTCA AACTGGTCGA GTTTACCGAC TATGTACGCC

```

-165-

251 CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTTCCAACAC
 301 AAACCTATC TTGACGACTT CAAAAAGAA CACAATCTGG ACATACCCGA
 351 AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
 401 AATCGCTGGA AGAAGTCAAA GACGGCAGCA CCGTATCCGC GCCCAACGAC
 451 CCGTCCAACT TCGCCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTGGAT
 501 CAAACTCAAA GACGGCATCA ATCCGTTGAC CGCATCCAAA GCGGACATCG
 551 CCGAGAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG
 601 CCGCGTAGCC GCGCCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
 651 CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
 701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
 751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
 801 GCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGAATGAAG
 851 GCGCAGCCAA ATAA

This corresponds to the amino acid sequence <SEQ ID 218; ORF4-1>:

1 MKTFFKTLA AALALILAAC GGQKDSAPAA SASAAADNGA AKKEIVFGTT
 51 VGDFGDMVKE IQAELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQH
 101 KPYLDDFKKE HNLDTIEVFO VPTAPLGLYP GKLSLEEVK DGSTVSAPND
 151 PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLKNI KIVELEAAQL
 201 PRSRADVDFV VVNGNYAISS GMKLTEALFO EPSFAYVNW AVKTADKDSQ
 251 WLKDVTEAYN SDAFKAYAHK RFEQYKSPAA WNEGAAK*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF4 shows 93.5% identity over a 93aa overlap with an ORF (ORF4a) from strain A of *N. meningitidis*:

25	orf4.pep	MKTFFKTLA AALALILAACG-QKDSAPAASASAAADNGA AKKEIVFGTTVGDFGDMVKE
	orf4a	MKTFFKTLA AALALILAACGGQKDSAPAASASAAADNGA AKKEIVFGTTVGDFGDMVKE
30	orf4.pep	QIQAELEKKGYTVKLVEFTDYVRPNLALAE GEL
	orf4a	XIQPELEKKGYTVKLVEFTDYVRPNLALAE GELDINXQHXXYLDXKXKXHNLDITXVXQ
35	orf4a	VPTAPLGLYPGKLSLXXVKXGSTVSAPNDPXXFXRVLVMDLXGIXKLKDXIXXXXXXX

The complete length ORF4a nucleotide sequence <SEQ ID 219> is:

40 1 ATGAAAACCT TCTTCAAAAC CCTTCCGCC GCGCACTCG CGCTCATCCT
 51 CGCCGCCTGC GCGGTCAAA AAGATAGCGC GCCCGCCGCA TCCGTTCTG
 101 CCGCCGCCGA CAACGCGCG GCGAANAAG AAATCGTCTT CGGCACGACC
 151 GTCGCGGACT TCGCGGATAT GGTCAAAGAA CANATCCAAC CCGAGCTGGA
 201 GAAAAAAGGC TACACCGTCA AACTGGTCTGA GTNTACCGAC TATGTGCGCN
 45 251 CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTTNCAACAC
 301 ANACNCTATC TTGACGACTN CAAAAANAA CACAATCTGG ACATCACCNN
 351 AGTCTTNCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
 401 AATCGCTGGA NNAAGTCAAA GANGGCAGCA CCGTATCCGC GCCCAACGAC
 451 CCGTNNNACT TCGNCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTNGAT
 50 501 CAAACTCAAA GACNGCATCA NNNNGNNGNN NNNANCNANA NNNGANANNN
 551 NNNNANNNT NNNNNNNNNN NNNNNCNCG NNNNNNNANN NNNNNNNNNN
 601 NCGNNTNNNN NNGCNNNNNT NNANNTNNN NNCNCCNNN NNNNTNNNN
 651 NANNANNAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
 701 TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
 55 751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
 801 CCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGAATGAAG
 851 GCGCAGCCAA ATAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 220>:

1 MKTFFKTLA AALALILAAC GGQKDSAPAA SASAAADNGA AKKEIVFGTT

51 VGDFGDMVKE XIQPELEKKG YTVKLVEFTD YVRXNLALAE GELDINVQKH
 101 XXYLDDXKKX HNLDITXVXQ VPTAPLGLYP GKLSLXXVK XGSTVSAPND
 151 PXXFXRVLVM LDELGXIKLK DXIXXXXXXX XXXXXXXXXX XXXXXXXXXX
 201 XXXXAXXXXX XXXXXXXXXX GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
 251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*

A leader peptide is underlined.

Further analysis of these strain A sequences revealed the complete DNA sequence <SEQ ID 221>:

1 ATGAAACCT TCTTCAAAC CCTTCCGCC GCCGACTCG CGCTCATCT
 51 CGCCGCTGC GCGGTCAAA AAGATAGCGC GCCGCGCGCA TCCGCTTCTG
 101 CCGCGCGCA CAACGGCGCG GCGAAAAAG AAATCGTCTT CGGCACGACC
 151 GTCGCGGACT TCGCGGATAT GGTCAAAGAA CAAATCCAAC CCGAGCTGGA
 201 GAAAAAGGC TACACGTC AACTGGTCA GTTACCGAC TATGTGCGCC
 251 CGAATCTGGC ATTGGCTGAG GCGGAGTTGG ACATCAACGT CTTCCAACAC
 301 AAACCTATC TTGACGACTT CAAAAAGAA CACAATCTGG ACATCACCGA
 351 AGTCTTCAA GTGCCGACCG CGCCTTTGGG ACTGTACCGG GCGAAGCTGA
 401 AATCGCTGA AGAAGTCAA GACGGCAGCA CCGTATCCGC GCCCAACGAC
 451 CCGTCCAAC TCGCCGCGT CTTGGTGATG CTCGACGAAC TGGGTGGAT
 501 CAACTCAA GACGGCATCA ATCCGCTGAC CGCATCCAAA GCGGACATTG
 551 CCGAAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCACTG
 601 CCGCGTAGCC GCGCGACGT GGATTTGCC GTCGTCAACG GCAACTACGC
 651 CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTCCAA GAACCGAGCT
 701 TTGCCATATG CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
 751 TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
 801 CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGAATGAAG
 851 GCGCAGCCAA ATAA

This encodes a protein having amino acid sequence <SEQ ID 222; ORF4a-1>:

1 MKTFFKTLA AALALILAAC GGQKDSAPAA SASAADNGA AKKEIVFGTT
 51 VGDFGDMVKE XIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINVQKH
 101 KPYLDDFKKE HNLDITEVFQ VPTAPLGLYP GKLSLEEVK DGSTVSAPND
 151 PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLKNI KIVELEAAQL
 201 PRSRADVFA VVNGNYAISS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
 251 WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*

ORF4a-1 and ORF4-1 show 99.7% identity in 287 aa overlap:

35	orf4a-1	10 20 30 40 50 60	MKTFFKTLA AALALILAAC GGQKDSAPAA SASAADNGA AKKEIVFGTT VGDFGDMVKE
	orf4-1	10 20 30 40 50 60	MKTFFKTLA AALALILAAC GGQKDSAPAA SASAADNGA AKKEIVFGTT VGDFGDMVKE
40	orf4a-1	70 80 90 100 110 120	QIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINVQKH KPYLDDFKKE HNLDITEVFQ
	orf4-1	70 80 90 100 110 120	QIQAELEKKG YTVKLVEFTD YVRPNLALAE GELDINVQKH KPYLDDFKKE HNLDITEVFQ
45	orf4a-1	130 140 150 160 170 180	VPTAPLGLYP GKLSLEEVK DGSTVSAPND PSNFARVLVM LDELGWIKLK DGINPLTASK
	orf4-1	130 140 150 160 170 180	VPTAPLGLYP GKLSLEEVK DGSTVSAPND PSNFARVLVM LDELGWIKLK DGINPLTASK
50	orf4a-1	190 200 210 220 230 240	ADIAENLKNI KIVELEAAQL PRSRADVFA VVNGNYAISS GMKLTEALFQ EPSFAYVNWS
	orf4-1	190 200 210 220 230 240	ADIAENLKNI KIVELEAAQL PRSRADVFA VVNGNYAISS GMKLTEALFQ EPSFAYVNWS
55	orf4a-1	250 260 270 280	AVKTADKDSQ WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAKX
	orf4-1	250 260 270 280	AVKTADKDSQ WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAKX

```

5      lip2.pasha          MNFKKLLGVALVSALALTACKDEKAQAP----
                                || | :|| || |:| | :| |
ORF4    VXTPNPDGRTPCPSLFFETATTSGENMKTFEKTLSAAAL--ALILAACGFKKTKARPPHPL
              110           120           130           140           150

10     lip2.pasha          30       40       50       60       70       80
-ATTAKTENKAPLKVGVMTGPEAQMTEVAVKIAEKYGLDVELVQFTEYTPQNAALHSKD
        : : | : | : | : || | |:| |:| |:| | | :
ORF4    LPPPTTARRRKEIVFGTTVGDFGDMVEQIQAELEKKGYTVKLVEFTDYVRNPLALAEGE
      160           170           180           190           200           210

15     lip2.pasha          90       100      110      120      130      140
LDANAFQTVPYLEQEVDKRGYKLAIIGNTLVWPPIAAYSKKIKNISLKDGTATVAIPNNAS
        |
ORF4    L.....

```

ORF4 shows 93.6% identity over a 94aa overlap with a predicted ORF (ORF4.ng) from *N. gonorrhoeae*:

```

25      orf4nm.pep      10      20      30
                        MKTFFKTL SAAALALILAACGXQKDSAPAA
                        |||||:::||||| |||||
      orf4ng      RANAVXTPNPDGRTPCLSFLFETATTSGENMKTFFKTLSTASLALILAACGGQKDSAPAA
                        200      210      220      230      240      250

30      orf4nm.pep      40      50      60      70      80      89
                        SASA-AADNGAAKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVFTDYVRPNLALA
                        ||:|:||||| ||||| ||||| ||||| ||||| |||||
      orf4ng      SAAAP SADNGAAKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVFTDYVRPNLALA
                        260      270      280      290      300      310

35      orf4nm.pep      90
                        EGEL
                        ||||
      orf4ng      EGELDINVQHKPYLDDFKKEHNL DITEAFQVPTAPLGLYPGKLKSLEEVKDGSTVSAPN
                        320      330      340      350      360      370

```

40 The complete length ORF4ng nucleotide sequence <SEQ ID 223> was predicted to encode a protein having amino acid sequence <SEQ ID 224>:

45

1	MKTFFKTLST	ASLALILAA	GGQKDSAPAA	SAAAPSADNG	AAKKEIVFGT
51	TVGDFGDMVK	EHQIAELEKK	GYYTVKLVEFT	DYVRPNLALA	EGELDINVFP
101	HKPYLDDEFK	EHNLDITEAF	QVPTAPLGLY	PGKLKSLEEV	KDGSTVSAPN
151	DPSNFARALV	MLNELGWIKL	KDGINPLTAS	KADIAENLKN	IKIVELEAAQ
201	LPRSRAADVDF	AVVNGNYAIS	SGMKLLEALF	QEPSFAYVNW	SAVKTADKDS
251	QWLKDVTEAY	NSDAFKAYAH	KRFEGYKYPA	AWNEGAAK*	

Further analysis revealed the complete length ORF4ng DNA sequence <SEQ ID 225> to be:

	1	atgAAAACCT	TCTTCAAAAC	cctttccgcc	gccgcaCTCG	CGCTCATCCT
50	51	CGCAGCCTGc	ggCggtcaAA	AAGACAGCGC	GCCCgcagcc	tctgcCGCCG
	101	CCCCTCTGc	CGATAACGgc	gCgCGGAAAA	AAGAAAtcgt	ctTCGGCAGC
	151	Accgtggggc	acttcggcgA	TatggtCAAA	GAACAATCC	AagcCGAgct
	201	gGAGAAAAAA	GgctACACcg	tcAAAttggt	cgaatttacc	gactatgtGC
	251	gCCCGAATCT	GGCATTGGCG	GAGGGCGAGT	TGGACATCAA	CGTCTTCCAA
55	301	CACAAACCTT	ATCTTGACGA	TTTCAAAAAA	GAACACAACC	TGGACATCAC
	351	CGAAGCCTTC	CAAAGTGCCGA	CCGCGCCTTT	GGGACTGTAT	CGGGGCAAAC
	401	TGAAATCGCT	GGAAGAAGTC	AAAGACGGCA	GCACCGTATC	CGCGCCCAac
	451	gACccgTCCA	ACTTCGCACG	CGCCTTGGTG	ATGCTGAACG	AACTGGGTTG
	501	GATCAAACTC	AAAGACGGCA	TCAATCCGCT	GACCGCATCC	AAAGCCGACA
60	551	TCGCGGAAAA	CTGAAAAAAC	ATCAAAATCG	TCGAGCTTGA	AGCCGCACAA

	1	<u>MKTFFKTL</u> <u>SA</u>	<u>AALALILA</u> <u>AC</u>	GGQKDSAPAA	SAAAPSADNG	AAKKEIVFGT
	51	TVGDFGDMVK	EQIQAELEKK	GYTVKLVEFT	DVVRPNLALA	EGELDINVFO
10	101	HKPYLDDEFK	EHNLDDITEAF	QVPTAPLGLY	PGKLSLEEV	KDGSTVSAPN
	151	DPSNFARALV	HMNELGWIKL	KDGINPLTAS	KADIAENLKN	IKIVELEAAQ
	201	LPRSRADVDF	AVVNGFYAIS	SGMKLTEALF	QNEPSAVVNW	SAVKTADKDS
	251	OWLKDVTEAY	NSDAFKAYAH	KRFEGYKYPA	AWEQGAAK*	

15			10	20	30	40	50	59
	orf4-1.pep	MKTFFKTL	SAALALILA	ACGGQKDS	SAPAASASA-	AADNGAAKKE	IVFGTTVG	DFGDMVK
	orf4ng-1	MKTFFKTL	SAALALILA	ACGGQKDS	SAPAASAA	PSADNGAAKKE	IVFGTTVG	DFGDMVK
20			10	20	30	40	50	60
	orf4-1.pep	60	70	80	90	100	110	119
	orf4ng-1	60	70	80	90	100	110	119
25			70	80	90	100	110	120
	orf4-1.pep	120	130	140	150	160	170	179
	orf4ng-1	120	130	140	150	160	170	179
30			130	140	150	160	170	180
	orf4-1.pep	180	190	200	210	220	230	239
	orf4ng-1	180	190	200	210	220	230	239
35			190	200	210	220	230	240
	orf4-1.pep	240	250	260	270	280		
	orf4ng-1	240	250	260	270	280		

[illegible]

-169-

	lip2_pasha	TEVAVKIAKEKYGLDVELVQFTEYTPQPNAAALHSKDLDANAFQTVPYLEQEVKDRGYKLAI	60	70	80	90	100	110
5	orf4ng-1.pep	AFQVPTAPLGLYPGKLKSLSEEVKDGSTVSAPNDPSNFARALVMLNELGWIKLKDGINPLT	120	130	140	150	160	170
	lip2_pasha	IGNTLVWPAAAYSKKIKNISELKDGATVAIPNNASNTARALLLQAHGLLKLKDPKN-VF	120	130	140	150	160	170
10	orf4ng-1.pep	ASKADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTE--ALFQEPSFA	180	190	200	210	220	230
	lip2_pasha	ATENDIENPKNIKIVQADTSLRLMLDDVELAVINNTYAGQAGLSPOKDGIIIVESKDSP	180	190	200	210	220	230
15	orf4ng-1.pep	YVNWSAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKYFAAWNEGAACK	240	250	260	270	280	289
20	lip2_pasha	YVNLVVSREDNKDDPRLQTFVKSFTTEVFQEAALKLENGGVVKGW	240	250	260	270		

Based on this analysis, including the homology with the outer membrane protein of *Pasteurella haemolítica*, and on the presence of a putative prokaryotic membrane lipoprotein lipid attachment site in the gonococcal protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF4-1 (30kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 8A and 8B show, respectively, the results of affinity purification of the His-fusion and GST-fusion proteins. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), Western blot (Figure 8C), FACS analysis (Figure 8D), and a bactericidal assay (Figure 8E). These experiments confirm that ORF4-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 8F shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF4-1.

Example 27

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 227>:

	1	CCTCGTCGTC	CTCGGCATGC	TCCAGTTTCA	AGGGGCGATT	TACTCCAAGG
	51	CGGTGGAACG	TATGCTCGGC	ACGGTCATCG	GGCTGGGCGC	GGGTTTGGGC
	101	GTTTTATGGC	TGAACACGCA	TTATTTCAC	GGCAACCTCC	TCTTCTACCT
40	151	CACCGTCGGC	ACGGCAAGCG	CACTGGCCGG	CTGGGCGGCG	GTCGGCAAAA
	201	ACGGCTACGT	CCCTmTGCTG	GCAGGGCTGA	CGATGTGTAT	GCTCATCGGC
	251	GACAACGGCA	GCGAATGGCT	CGACAGCGGA	CTCATGCGCG	CCATGAACGT
	301	CCTCATCGGC	GyGGCCATCG	CCATCGCCGC	CGCCAACTG	CTGCCGCTGA
	351	AATCCACACT	GATGTGGCGT	TTCATGCTTG	CCGACAACCT	GGCCGACTGC
45	401	AGCAAAATGA	TGCGCGAAAT	CAGCAACGGC	AGGCGCATGA	CCCGCGAAGC
	451	CCTCGAGGAG	AACATGGCGA	AAATGCGCCA	AATCAACGCA	CGCATGGTCA
	501	AAAGCCGCG	CCATCTCGCC	GCCACATCGG	GCGAAAGCTG	CATCAGCCCC
	551	GCCATGATGG	AAGCCATGCA	GCACGCCAC	CGTAAATCG	TCAACACCAC
	601	CGAGCTGCTC	CTGACCACCG	CCGCCAAGCT	GCAATCTCCC	AAACTCAACG

5

This corresponds to the amino acid sequence <SEQ ID 228; ORF8>:

10

15 Computer analysis of this amino acid sequence gave the following results:

Sequence motifs

ORF8 is proline-rich and has a distribution of proline residues consistent with a surface localization. Furthermore the presence of an RGD motif may indicate a possible role in bacterial adhesion events.

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF8 shows 86.5% identity over a 312aa overlap with a predicted ORF (ORF8.ng) from *N. gonorrhoeae*:

25

50 The complete length ORF8ng nucleotide sequence <SEQ ID 229> is predicted to encode a protein having amino acid sequence <SEQ ID 230>:

55

201 QNRQHHRAAP DHRROAAISQ TQRQNPAAAR PPLHTAPNRP ATNRRPHQRQ
 251 TRPPHPRHR HQPRTGSPRR TPPLPMAGFP LAQHQYASGN FRPRHPPATH
 301 PPQMACPRT PTPAPKPA*

Based on the sequence motifs in these proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 28

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 231>:

```

10      1  . . GAAATCAGCC TGC GG TCCGA CNACAGGCCG GTTCCGTGN CGAAGCGGCG
      51  GGATTCGGAA CGTTTTCTGC TGTGGACGG CGGCAACAGC CGGCTCAAGT
      101  GGGCGTGGGT GGA AACCGC AC GTTCGCAA CCGTCGGTAG CGCGCCGTAC
      151  CGCGATTGT CGCCTTTGGG CGCGGAGTGG GCGGAAAGG CGGATGGAAA
      201  TGTC CGCATC GTCGGTTGCG CTGTGTGCGG AGAATTCAAA AAGGCACAAG
      251  TGCAGGAACA GTCGCCCCGA AAAATCGAGT GGCTGCCGTC TTCCGCACAG
      301  GCTTT .GGCA TACGCAACCA CTACCGCCAC CCCGAAGAAC ACGGTTCCGA
      351  CCGCTGGTTC AACGCCTTGG GCAGCCGCCG CTTACGCCG AACGCCTGCG
      401  TCGTCGTCAG TTGCGGCACG GCGGTAACGG TTGACGCGCT CACCGATGAC
      451  GGACATTATC TCGGAGA .GG AACCATCATG CCCGTTTCC ACCTGATGAA
      501  AGAATCGCTC GCCGTCCGAA CCGCCAACCT CAACCGGCAC GCCGGTAAGC
      551  GTTATCCTTT CCCGACCGG. .
  
```

This corresponds to the amino acid sequence <SEQ ID 232; ORF61>:

```

      1  . . EISLRSDXRP VSVXKRDRSE RFLLLDGGNS RLKAWVENG TFATVGSAPY
      51  RDLSPGLAEW AEKADGNVRI VGCAVCGEFK KAQVQEQRLAR KIEWLPSSAQ
      101  AXGIRNHYRH PEEHSGDRWF NALGSRREFSR NACVVVSCGT AVTVDALTD
      151  GHYLGXGTIM PGFHLMKESL AVRTANLNRH AGKRYPFPT. .
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 233>:

```

      1  ATGACGGTTT TGAAGCTTTC GCACTGGCGS GTGTTGGCGG AGCTTGCCGA
      51  CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
      101  CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATACGCGGG
      151  CTGTTGCGCC AACACGACGG CTATTGGCGG CTGTTGCGCC CATTGGCGGT
      201  TTTCGATGCC GAAGGTTTGC GCGAGCTGGG GGAAGGTCG GGTTTTCAGA
      251  CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
      301  GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGCG TGACCCACCT
      351  GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
      401  GCGAGTGTCT GATGTTTCA GTTGGCTGGG TGTGTGACCG GCCGCAGTAT
      451  GAGTTGGGTT CGCTGTCGCC TGTGCGGCA GTGGCGTGTG GCGCGCCCTT
      501  GTCGCGTTTA GGTGTTGATG TGCAGATTAA GTGGCCCAAT GATTGTTGTT
      551  TCGGACCGCA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC
      601  GGCAAAACGG TTGCCGTGGT CCGTATCGGC ATCAATTTTG TCCTGCCCAA
      651  GGAAGTAGAA AATGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC
      701  GGCGGGGCAA TGCCGATGCC GCCGTGCTGC TGGAAACGCT GTTGTGGGAA
      751  CTGGACGCGG TGTGTTGCA ATATGCGCGG GACGGATTG CGCCTTTTGT
      801  GGCGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
      851  TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGGACGGA
      901  CAAGGCGTTT TGCATTGGA AACGGCAGAG GGCACACAGA CGGTCGTCAG
      951  CGGCGAAATC AGCCTGCGGT CCGACGACAG GCCGTTTCC GTGCCGAAGC
      1001  GCGCGGATTC GGAACGTTTT CTGCTGTTGG ACGGCGGCAA CAGCCGCTC
      1051  AAGTGGGCGT GGGTGGAAAA CGGCACGTTC GCAACCGTCG GTAGCGCGCC
      1101  GTACCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
      1151  GAAATGTCCG CATCGTCGGT TGCGCTGTGT GCGGAGAATT CAAAAAGGCA
      1201  CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGSCTGC CGTCTTCCGC
      1251  ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCGAA GAACACGGTT
      1301  CCGACCGCTG GTTCAACGCC TTGGGCAGCC GCCGTTTCA CCGCAACGCC
      1351  TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
      1401  TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
      1451  AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
      1501  CGTTATCCTT TCCCGACCAC AACGGGCAAT GCCGTCGCCA GCGGCATGAT
      1551  GGATGCGGTT TCGGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA
      1601  AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
  
```

```

1651 GCAAAAGTTG CCGAAGCCCT GCCGCCTGCA TTTTGGCGG AAAATACCGT
1701 GCGCGTGGCG GACAACCTCG TCATTACGG GTTGTGAAC ATGATTGCCG
1751 CCGAAGGCAG GGAATATGAA CATATTAA

```

This corresponds to the amino acid sequence <SEQ ID 234; ORF61-1>:

```

5      1  MTLVKLSHR  VLAELADGLP  QHVSQALRMA  DMKPQQLNGF  WQQMPAHIRG
      51  LLRQHDGYWR  LVRPLAVFDA  EGLRELGERS  GFQTALKHEC  ASSNDEILEL
     101  ARIAPDKAHK  TICVTHLQSK  GRGRQGRKWS  HRLGECLMFS  FGWVFDPRQY
     151  ELGSLSPVAA  VACRRALSRL  GLDVQIKWPN  DLVVGRDKLG  GILIETVRTG
     201  GKTVAVVGIG  INFVLPEKEV  NAASVQSLFQ  TASRRGNADA  AVLLETLLVE
     251  LDAVLLQYAR  DGFAPFVAEY  QANRDHGKA  VLLLRDGETV  FEGTVKGVDPG
     301  QGVLHLETAE  GKQTVVSGEI  SLRSDDRPVS  VPKRRDSERF  LLLDGGNSRL
     351  KWAUVENGTF  ATVGSAPYRD  LSPLGAEWAE  KADGNVRIVG  CAVCGEFKKA
     401  QVQEQLARKI  EWLPSSAQAL  GIRNHYRHPE  EHGSDRWFNA  LGSRRFRSNA
     451  CVVVSCTAV  TVDALTDGTH  YLGGTIMPGF  HLMKESLAVR  TANLNRHAGK
    15  501  RYPFPTTGN  AVASGMMDAV  CGSVMMHGR  LKEKTGAGKP  VDVIIITGGGA
      551  AKVAEALPPA  FLAENTVRVA  DNLVIYGLLN  MIAAEGREYE  HI*

```

Figure 9 shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF61-1. Further computer analysis of this amino acid sequence gave the following results:

Homology with the baf protein of *B. pertussis* (accession number U12020).

20 ORF61 and baf protein show 33% aa identity in 166aa overlap:

```

      orf61 23  LLLDGGNSRLKWAWE-NGTFATVGSAPYR----DLSPGLAEWAEKADGNVRIVGCAVCG 77
      +L+D GNSRLK W + + A AP DL LG A R +G V G
      baf 3  ILIDSGNSRLKVGWFDPAQAAREPAPVAFDNLDLALGRWLATLPRRPQALGVNVAG 62
    25  orf61 78  EFKAQVQEQLAR---KIEWLPSSAQAXGIRNHYRHPEEHGSDRW---FNLGSRFRFSRN 131
      + + L I WL + A G+RN YR+P++ G+DRW L +
      baf 63  LARGEIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDLGADRWACMVGLARQPSVHP 122
    30  orf61 132 ACVVVSCGTAVTVDALTDGTHYLXGTIMPGFHLMKESLAVRTANL 177
      +V S GTA T+D + D + G G I+PG +M+ +LA TA+L
      baf 123 PLLVASFGTATTLDITGPDNVFPG-GLILPGPAMMRGALAYGTAHL 167

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF61 shows 97.4% identity over a 189aa overlap with an ORF (ORF61a) from strain A of *N.*

35 *meningitidis*:

```

      orf61.pep
      orf61a  TVFEGTVKGVDPGQGVHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNS
    40      290      300      310      320      330      340
      orf61.pep  RLKWAUVENGTFATVGSAPYRDLSPGLAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLAR
      orf61a  RLKWAUVENGTFATVGSAPYRDLSPGLAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLAR
    45      350      360      370      380      390      400
      orf61.pep  KIEWLPSSAQAXGIRNHYRHPEEHGSDRWFNALGSRFRFSRNACVVVSCGTAVTVDALTD
      orf61a  KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRFRFSRNACVVVSCGTAVTVDALTD
    50      410      420      430      440      450      460
      orf61.pep  GHYLGXGTIMPGFHLMKESLAVRTANLNRHAGKRYFPPT
      orf61a  GHYLG-GTIMPGFHLMKESLAVRTANLNRHAGKRYFPPTTGNASGMMDAVCGSVMM
    55      470      480      490      500      510      520
      orf61a  HGRLKEKTGAGKPDVVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGG
    60

```

530 540 550 560 570 580

The complete length ORF61a nucleotide sequence <SEQ ID 235> is:

```

1  ATGACGGTTT TGAAGCCTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA
5  51  CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
101 CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATACGCGGG
151 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTTGCGGT
201 TTTGATGCC GAAGGTTTGC GCGAGCTGGG GGAAAGGTCG GGTTTTCAGA
251 CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
301 GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGTG TGACCCACCT
10 351 GCAAAGTAAG GGCAGGGGCG GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
401 GCGAGTGTCT GATGTTTCACT TTTGGCTGGG TGTGTTGACCG GCCGCAGTAT
451 GAGTTGGGTT CGCTGTCGCC TGTGCGGCA GTGGCGTGCC GGCGCGCCTT
501 GTCGCGTTTG GGTGTTGAAA CGCAAATCAA GTGGCCAAAC GATTGTCGTC
15 551 TCGGACGCGA CAAATTGGGC GGCAATCTGA TTGAAACGGT CAGGACGGGC
601 GGCAAAACGG TTGCCGTGGT CCGTATCGGC ATCAATTTCTG TGCTGCCCAA
651 GGAAGTGGAA AACGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC
701 GCGGGGGAAA TGCCGATGCC GCCGTGTTGC TGGAAACGCT GTTGGCGGAA
751 CTTGATGCGG TGTGTTGCA ATATGCGCGG GACGGATTG CGCCTTTTGT
801 GCGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
20 851 TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGGACGGA
901 CAAGGCGTTC TGCACTTGA AACGGCAGAG GGCAAAACAGA CCGTCGTCAG
951 CCGCGAAATC AGCCTGCGGT CCGACGACAG GCCGTTTTCC GTGCCGAAGC
1001 GCGCGGATTG GGAACGTTTT CTGCTGTTGG ACGCGGGCAA CAGCCGGCTC
1051 AAGTGGGCGT GGGTGAAAA CGGCACGTTT GCAACCGTCG GTAGCGCGCC
25 1101 GTACCGCGAT TTGTCGCTT TGGGCGCGGA GTGGGCGGAA AAGGTGGATG
1151 GAAATGTCCG CATCGTCGGT TGGCGCGTGT GCGGAGAATT CAAAAAGGCA
1201 CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCGAA GAACACGGTT
1301 CCGACCGCTG GTTCAACGCC TTGGGCAGCC GCCGCTTCAG CCGCAACGCC
30 1351 TGCGTCGTCG TCAGTTGCGG CACGCGGTA ACGGTTGACG CGCTCACCGA
1401 TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
1451 AAGAATCGCT CGCCGTCGGA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
1501 CGTTATCCTT TCCCAGCCAC AACGGGCAAT GCCGTCGCCA GCGGCATGAT
1551 GGATGCGGTT TCGGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA
35 1601 AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
1651 GCAAAAGTTG CCGAAGCCCT GCCGCTGCA TTTTGGCGG AAAATACCGT
1701 GCGCGTGGCG GACAACCTCG TCATTACCGG GCTGCTGAAC CTGATTGCCG
1751 CCGAAGGCGG GGAATCGGAA CATACTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 236>:

```

40 1  MTVLKPSHWR VLAELADGLP QHVSQ LARMA DMKPQQLNGF WQOMPAHIRG
51  LLRQHDGYWR LVRPLAVFDA EGLREL GERS GFQTALKHEC ASSNDEILEL
101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWVDRPOY
151 ELGSLSPVAA VACRRALSRL GLKTQIKWPN DLVVGRDKLG GILIETVRTG
201 GKTAVVVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE
45 251 LDAVLLQYAR DGFAPFVAEY QANRDHGKA VLLLRDGETV FEGTVKGV DG
301 QGV LHLETA E GKQTVVSGEI SLRSDDRPVS VPKRRDSERF LLLDGGNSRL
351 KWAUVENGTF ATVGSAPYRD LSPLGAEWAE KVDGNVRIVG CAVCGEFKKA
401 QVQEQLARKI EWLPSSAQAL GIRNHYRHPE EHGS DRWFNA LGSRRF SRNA
451 CVV VSCGTAV TVDALTD DGH YLGGT IMPGF HLMKESLAVR TANLNRHAGK
50 501 RYFPFTTGN AVASGMDAV CGSVMMHGR LKEKTGAGKP VDVIITGGGA
551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HT*

```

ORF61a and ORF61-1 show 98.5% identity in 591 aa overlap:

```

55 orf61a.pep 10 20 30 40 50 60
    MTVLKPSHWRVLAELADGLPQHVSQ LARMADMKPQQLNGFWQOMPAHIRGLLRQHDGYWR
    orf61-1 10 20 30 40 50 60
    MTVLKLSHWRVLAELADGLPQHVSQ LARMADMKPQQLNGFWQOMPAHIRGLLRQHDGYWR

60 orf61a.pep 70 80 90 100 110 120
    LVRPLAVFDAEGLREL GERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK
    orf61-1 70 80 90 100 110 120
    LVRPLAVFDAEGLREL GERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK

65 130 140 150 160 170 180

```

5	orf61a.pep	GRGRQGRKWSHRLGECLMFSFGWVFD RPQYELGSLSPVAACRRALSRLGLKTQIKWPN	
	orf61-1	GRGRQGRKWSHRLGECLMFSFGWVFD RPQYELGSLSPVAACRRALSRLGLDVQIKWPN	130 140 150 160 170 180
10	orf61a.pep	DLVVGRDKLGGIL IETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	190 200 210 220 230 240
	orf61-1	DLVVGRDKLGGIL IETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	190 200 210 220 230 240
15	orf61a.pep	AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGV D G	250 260 270 280 290 300
	orf61-1	AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGV D G	250 260 270 280 290 300
20	orf61a.pep	QGV L H L E T A E G K Q T V V S G E I S L R S D D R P V S V P K R D S E R F L L D G G N S R L K W A W E N G T F	310 320 330 340 350 360
	orf61-1	QGV L H L E T A E G K Q T V V S G E I S L R S D D R P V S V P K R D S E R F L L D G G N S R L K W A W E N G T F	310 320 330 340 350 360
25	orf61a.pep	A T V G S A P Y R D L S P L G A E W A E K D G N V R I V G C A V C G E F K K A Q V Q E Q L A R K I E W L P S S A Q A L	370 380 390 400 410 420
	orf61-1	A T V G S A P Y R D L S P L G A E W A E K A D G N V R I V G C A V C G E F K K A Q V Q E Q L A R K I E W L P S S A Q A L	370 380 390 400 410 420
30	orf61a.pep	G I R N H Y R H P E E H G S D R W F N A L G S R R F S R N A C V V S C G T A V T V D A L T D D G H Y L G G T I M P G F	430 440 450 460 470 480
	orf61-1	G I R N H Y R H P E E H G S D R W F N A L G S R R F S R N A C V V S C G T A V T V D A L T D D G H Y L G G T I M P G F	430 440 450 460 470 480
35	orf61a.pep	H L M K E S L A V R T A N L N R H A G K R Y P F P T T T G N A V A S G M M D A V C G S V M M H G R L K E K T G A G K P	490 500 510 520 530 540
	orf61-1	H L M K E S L A V R T A N L N R H A G K R Y P F P T T T G N A V A S G M M D A V C G S V M M H G R L K E K T G A G K P	490 500 510 520 530 540
40	orf61a.pep	V D V I I T G G G A A K V A E A L P P A F L A E N T V R V A D N L V I H G L L N L I A A E G G E S E H T X	550 560 570 580 590
	orf61-1	V D V I I T G G G A A K V A E A L P P A F L A E N T V R V A D N L V I Y G L L N M I A A E G G E Y E H I X	550 560 570 580 590

Homology with a predicted ORF from *N.gonorrhoeae*

ORF61 shows 94.2% identity over a 189aa overlap with a predicted ORF (ORF61.ng) from *N.*

50 *gonorrhoeae*:

55	orf61.pep	E I S L R S D X R P V S V X K R R D S E R F L L D G G N S	30
	orf61ng	T V C E G T V K G V D G R G V L H L E T A E G E Q T V V S G E I S L R P D N R S V S V P K R P D S E R F L L E G G N S	211
60	orf61.pep	R L K W A W E N G T F A T V G S A P Y R D L S P L G A E W A E K A D G N V R I V G C A V C G E F K K A Q V Q E Q L A R	90
	orf61ng	R L K W A W E N G T F A T V G S A P Y R D L S P L G A E W A E K A D G N V R I V G C A V C G S K K A Q V Q E Q L A R	271
65	orf61.pep	K I E W L P S S A Q A X G I R N H Y R H P E E H G S D R W F N A L G S R R F S R N A C V V S C G T A V T V D A L T D D	150
	orf61ng	K I E W L P S S A Q A L G I R N H Y R H P E E H G S D R W F N A L G S R R F S R N A C V V S C G T A V T V D A L T D D	331
65	orf61.pep	G H Y L G X G T I M P G F H L M K E S L A V R T A N L N R H A G K R Y P F P T	189
	orf61ng	G H Y L G - G T I M P G F H L M K E S L A V R T A N L N R P A G K R Y P F P T T T G N A V A S G M M D A V C G S I M M	390

An ORF61ng nucleotide sequence <SEQ ID 237> was predicted to encode a protein having amino acid sequence <SEQ ID 238>:

```

1  MFSFGWAFDR PQYELGSLSP VAALACRRAL GCLGLETQIK WPNDLVVGRD
51 KLGGLIETV RAGGKTVAUV GIGINFVLPK EVENAASVQS LFQTASRRGN
101 ADAAVLLET LLAELGAVLEQ YAEFGFAPFL NEYETANRDH GKAVLLLRDG
151 ETVCEGTVKG VDGRGVLHLE TAEGEQTVVS GEISLRPDNR SVSVFKRPDS
201 ERFLLLEGNN SRLKWAVVEN GTFATVGSAP YRDLSPLGAE WAEKADGNVR
251 IVGCAVCGES KKAQVKEOLA RKIEWLPSSA QALGIRNHYR HPEEHGSDRW
301 FNALGSRRFS RNACVVVSCG TAVTVDALTD DGHYLGGTIM PGFHLMKESL
10 351 AVRTANLNRP AGKRYPFPTT TGNVASGMM DAVCGSIMMM HGRLEKKNKA
401 GKPVDVIITG GGAAKVAEAL PPAFLAENTV RVADNLVIHG LLNLIAAEGG
451 ESEHA*

```

Further analysis revealed the complete gonococcal DNA sequence <SEQ ID 239> to be:

```

1  ATGACGGTTT TGAAGCCTTC GCATTGGCGG GTGTTGGCGG AGCTTGCCGA
15 51 CGGTTTGCCG CAACACGTAT CGCAATTGGC GCGTGAGGCG GACATGAAGC
101 CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA TATACGCGGG
151 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CCTTTGGCGT
201 TTTCGATGCC GAAGGTTTGC GCGATCTGGG GGAAAGGTCG GGTTTTCAGA
251 CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
301 GCGCGGATTG CGCCGGACAA GCGGCACAAA ACCATATGCG TGACCCACCT
351 GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGCTCG CACCGTTTGG
401 GCGAGTGCCT GATGTTCACT TTCGGCTGGG CGTTTGACCG GCCGCAGTAT
451 GAGTTGGGTT CGCTGTCGCC TGTTGCGGCA CTTGCGTGCC GCGCGCTTTT
501 GGGGTGTTTG GGTTTGAAA CGCAAATCAA GTGGCCAAAC GATTTGGTCG
25 551 TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACAGT CAGGGCGGGC
601 GGTAAAACGG TTGCCGTGGT CCGTATCGGC ATCAATTTCT TGCTGCCCAA
651 GGAAGTGGA AAGCGCGCTT CCGTGAGTCT GCTGTTTCAG ACGGCATCGC
701 GCGCGGGCAA TGCCGATGCC GCCGTATTGC TGGAACATT GCTTGCGGAA
30 751 CTGGGCGCGG TGTGGAACA ATATGCGGAA GAAGGGTTCT CGCCATTTT
801 AAATGAGTAT GAAACGGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
851 TCGCGCAGCG GAAACCGGTG TCGCAAGGCA CGGTTAAAGG CGTGACGGA
901 CGAGGCGTTC TGCACCTTGA AACGGCAGaa ggcgaACAGa cggtcgtcag
951 cggcgaaaTC AGcctGCggc ccgacaacaG GTCGGtttcc gtgccgaagc
1001 ggcgggatTC GgaacgtTTT tTGctgttgg aaggcgggaa cagccgGCTC
35 1051 AAGTGGGCGT GggtggAAAa cggcacgttc gcaaccgtgg gcagcgCgC
1101 gtaCCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
1151 GAAATGTCCG CATCGTCGGT TCGCGCGTGT GCGGAGAATC CAAAAGGCA
1201 CAAGTGAAGG AACAGCTCGC CCGAAAATC GAGTGGCTGC CGTCTTCCCG
1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
40 1301 CCGACCGTTG GTTCAACGCC TTGGGCGAGC GCGCTTCAG CCGCAACGCC
1351 TCGCTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
1401 TGACGGACAT TATCTCGCGG GAACCATCAT GCCCGGCTTC CACCTGATGA
1451 AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGCCC CGCCGGCAAA
1501 CGTTACCCTT TCCCGACCAC AACGGGCAAC GCCGTGCGAA CGGCGATGAT
45 1551 GGACGCGGTT TCGGCTCGA TAATGATGAT GCACGGCCGT TTGAAAGAAA
1601 AAAACGGCGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
1651 GCGAAAGTCG CCGAAGCCCT GCCGCTGCA TTTTGGCGG AAAATACCGT
1701 GCGCGTGGCG GACAACCTCG TCATCCACGG GCTGCTGAAC CTGATTGCCG
1751 CCGAAGGCGG GGAATCGGAA CACGCTTAA

```

50 This corresponds to the amino acid sequence <SEQ ID 240; ORF61ng-1>:

```

1  MTVLKPSHWR VLAELADGLP QHVSQALAREA DMKPQQLNGF WQMPAHIRG
51 LLRQHDGYWR LVRPLAVFDA EGLRDLGERS GFQTALKHEC ASSNDEILEL
101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWAFDRPOY
55 151 ELGSLSPVAA LACRRALGCL GLETQIKWPN DLVVGRDKLG GILLETVRAG
201 GKTVAUVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE
251 LGAVLEQYAE EGFAPFLNEY ETANRDHGKA VLLLRDGETV CEGTVKGVDG
301 RGVLEHLETAE GEQTVVSGEI SLRPDNRSVS VPKRPDSERF LLLEGNSRL
351 KWAVVENGTF ATVGSAPYRD LSPLGAEWAE KADGNVRIWG CAVCGESKKA
401 QVKEQLARKI EWLPSAQAAL GIRNHYRHPE EHGSDRWANA LGSRRFSRNA
60 451 CVVVSCGTAV TVDALTDGHH YLGGTIMPGF HLMKESLAVR TANLNRPAKG
501 RYPFPTTTGN AVASGMMDAV CGSIMMMHGR LKEKNAGAGP VDVITGGGA
551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HA*

```

ORF61ng-1 and ORF61-1 show 93.9% identity in 591 aa overlap:

	orf61ng-1.pep	MTVLKPSHWRVLAELADGLPQHVSQALAREADMKPQQLNGFWQOMPAHIRGLLRQHDGYWR	60
	orf61-1	MTVLKLSHWRVLAELADGLPQHVSQALARMADMKPQQLNGFWQOMPAHIRGLLRQHDGYWR	60
5	orf61ng-1.pep	LVRPLAVFDAEGLRDLGERSGFTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK	120
	orf61-1	LVRPLAVFDAEGLRELGERSGFTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK	120
10	orf61ng-1.pep	GRGRQGRKWSHRLGECIMFSGFWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPN	180
	orf61-1	GRGRQGRKWSHRLGECIMFSGWVDFRQYELGSLSPVAACRRALSRLGLDVQIKWPN	180
15	orf61ng-1.pep	DLVVGRDKLGGILIIETVRAGGKTAVVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	240
	orf61-1	DLVVGRDKLGGILIIETVRTGGKTAVVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA	240
20	orf61ng-1.pep	AVLLETLLAELGAVLEQYAEAGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVGD	300
	orf61-1	AVLLETLLVELDAVLLQYARDGFAPFAEYQAANRDHGKAVLLLRDGETVCEGTVKGVGD	300
25	orf61ng-1.pep	RGVLHLETAEGEQTVVSGEISLRPNRSVSPKRPDSEFLLLEGNSRLKWAWVENGTF	360
	orf61-1	QGVLEHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSEFLLLDGNSRLKWAWVENGTF	360
30	orf61ng-1.pep	ATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQAL	420
	orf61-1	ATVGSAPYRDLSPGLAEWAEEKADGNVRIVGCAVCGEFKKAQVQEQQLARKIEWLPSSAQAL	420
35	orf61ng-1.pep	GIRNHYRHPEEHGSDRWFNALGSRFRSRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF	480
	orf61-1	GIRNHYRHPEEHGSDRWFNALGSRFRSRNACVVVSCGTAVTVDALTDGHHYLGGTIMPGF	480
40	orf61ng-1.pep	HLMKESLAVRTANLNRPAKGKRYPFPTTTGNAVASGMMDAVCGSIMMHGRLKEKNAGKGP	540
	orf61-1	HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKGP	540
	orf61ng-1.pep	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLI AEGGESEHAX	593
	orf61-1	VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMI AEGGREYEHIX	593

Based on this analysis, including the homology with the baf protein of *B.pertussis* and the presence of a putative prokaryotic membrane lipoprotein lipid attachment site, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

45 Example 29

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 241>:

	1	ATGTTTACC	AAATCCTTGC	CCTGATTATC	TGGAGCAGCT	CGTTTATTGC
	51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCGGCGTGC
	101	GCCTGCTAAT	TGCCGCGCTG	CCTGCACTGC	CCGCTGCCC	CGTCATGTC
50	151	GGCAAGATTC	CGCGTAGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTCGT
	201	CAACTATGTG	CTGACCCTGC	TGCTTCAGTT	TGTCGGGTTG	AAATACACTT
	251	CCGCCGCCAG	CGCATCGGTC	ATTGTCGGAC	TCGAGCCGCT	GCTGATGGTG
	301	TTTGTCGGAC	ACTTTTCTT	CAACGACAAA	GCGCGTGCCT	ACCACTGGAT
	351	ATGCGGCGCG	GCGGCATTG	CCGGTGTGCG	GCTGCTGATG	GCGGGCGGTG
55	401	CGGAAGAGGG	CGGCGAAGTC	GGCTGGTTCG	GCTGCCTGCT	GGTGTGTGTTG
	451	GCGGGCGCGG	GCTTTTGTGC	CGTATGCGT	CCGACGCAAA	GGCTGATTGC
	501	ACGCATCGGC	GCACCGGCAT	TCACATCTGT	TTCCATTGCC	GCCGCATCGT
	551	TGATGTGCCT	GCGTTTTTCG	CTTGCTTTGG	CGCAAAGTTA	TACCGTGGAC
	601	TGGAGCGTCG	GGATGGTATT	GTCGCTGCTG	TATTTGGGTT	TGGGGTGC..

60 This corresponds to the amino acid sequence <SEQ ID 242; ORF62>:

1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLIIAAL PALPACRRHV
 51 GKIPREEWKP LLIVSFVNYV LTLLQFVGL KYTSAASASV IVGLEPLLMV
 101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
 151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
 201 WSVGMVLSLL YLGLGC..

Further work revealed the complete nucleotide sequence <SEQ ID 243>:

1 ATGTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTATTGC
 51 CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG STCGGCGTGC
 101 GCCTGCTAAT TGCCGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC
 151 GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTCTG
 201 CAACTATGTG CTGACCCTGC TGCTTCAGTT TGTCGGGTTG AAATACACTT
 251 CCGCCGCCAG CGCATCGGTC ATTGTGCGAC TCGAGCCGCT GCTGATGGTG
 301 TTTGTGCGAC ACTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
 351 ATGCGGCGCG GCGGCATTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
 401 CGGAAGAGGG CGGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTG
 451 GCGGGCGCGG GCTTTTGTGC CGCTATGCGT CCGACGCAA GGCTGATTGC
 501 ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT
 551 TGATGTGCCT GCCGTTTTCG CTGCTTTGG CGCAAAGTTA TACCGTGGAC
 601 TGGAGCGTCG GGATGGTATT GTCGCTGCTG TATTTGGGTT TGGGGTGCAG
 651 CTGGTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCTGCCA
 701 ATGTTTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCGG CGTGCTGCTG
 751 GCGGTTTTGA TTTTGGGCGA ACACCTGTCG CCCGTGTCCG CCTTGGGCGT
 801 GTTGTGCTG ATCGCCGCCA CCTTGGTTCG CGGCCGGCTG TCGCATCAA
 851 AATAA

25 This corresponds to the amino acid sequence <SEQ ID 244; ORF62-1>:

1 MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLIIAAL PALPACRRHV
 51 GKIPREEWKP LLIVSFVNYV LTLLQFVGL KYTSAASASV IVGLEPLLMV
 101 FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
 151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
 201 WSVGMVLSLL YLGLGCGWYA YWLWNKMSR VPANVSGLLI SLEPVVGVL
 251 AVLILGEHLS PVSALGVFVV IAATLVAGRL SHQK*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical transmembrane protein HI0976 of *H. influenzae* (accession number Q57147)

ORF62 and HI0976 show 50% aa identity in 114aa overlap:

35 Orf62 1 MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHV
 M YQILAL+IWSSS I K Y +D? L+V VR R KI + K
 HI0976 1 MLYQILALLIWSSSLIVGKLTYSMMDPVLVQVRLIIAMIVMPLFLRRWKIDKPMRKQ 60
 40 Orf62 61 LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLMVFGHHFFNDKARAY 114
 L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VFVGHFF K +
 HI0976 61 LWWLAFFNYTAVFLLQFVGLKYTSAASAVTMIGLEPLLVVFGHHFFKTKQNGF 114

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF62 shows 99.5% identity over a 216aa overlap with an ORF (ORF62a) from strain A of *N.*

45 *meningitidis*:

10 20 30 40 50 60
 orf62.pep MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHV
 orf62a MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLIIAALPALPACRRHV
 50 10 20 30 40 50 60
 70 80 90 100 110 120
 orf62.pep LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLMVFGHHFFNDKARAYHWICGA
 orf62a LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLMVFGHHFFNDKARAYHWICGA
 55 70 80 90 100 110 120
 130 140 150 160 170 180
 orf62.pep AAFAGVALLMAGGAEEGGEVGFCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA

-178-

```

|||||
orf62a  AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA
          130      140      150      160      170      180

5      190      200      210
orf62.pep AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC
          |||||
orf62a  AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI
          190      200      210      220      230      240

10     orf62a  SLEPVVGVLAVLILGEHLSPVSVLGVFVVIAATLVAGRLSHQKX
          250      260      270      280

```

The complete length ORF62a nucleotide sequence <SEQ ID 245> is:

```

15      1  ATGTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTATTGC
51      51  CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC
101     101  GCCTGCTGAT TGCTGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC
151     151  GGCAAGATTC CCGGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTTCGT
201     201  CAACTATGTG CTGACCCTGC TACTTCAGTT TGTGCGGTTG AAATACACTT
251     251  CCGCCGCCAG CGCATCGGTC ATTGTCGGAC TCGAGCCACT GCTGATGGTG
20      301  TTTGTCGGAC ACTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
351     351  ATGCGGCGCG GCGGCATTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
401     401  CGGAAGAGGG CGGCGAAGTC GGCTGGTTCC GCTGCTGCTT GGTGTTGTTG
451     451  GCGGGCGCGG GCTTTGTGTC CGCTATGCGT CCGACGCAAA GGCTGATTGC
501     501  ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT
25      551  TGATGTGCCT GCCGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC
601     601  TGGAGCGTCG GAATGGTATT GTCGCTGCTG TATTGGGCGG TGGGGTGCAG
651     651  CTGGTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCCTGCCA
701     701  ACGTTTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCGG CGTGTGCTG
751     751  GCGGTTTGA TTTTGGGCGA ACACCTGTCG CCCGTGTCCG TCTTGGGCGT
30      801  GTTGTGCTC ATCGCCGCCA CCTTGGTTGC CCGCCGCGTG TCGCATCAA
851     851  AATAA

```

This encodes a protein having amino acid sequence <SEQ ID 246>:

```

35      1  MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLLIAAL PALPACRRHV
51      51  GKIPREEWKP LLIVSFVNYV LTLQLQFVGL KYTSAASASV IVGLEPLLMV
101     101  FVGHHFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGLLVLL
151     151  AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
201     201  WSVGMVLSLL YLGVCWSYAY YWLWNKGMSR VPANVSGLLI SLEPVVGVL
251     251  AVLILGEHLS PVSVLGVFVV IAATLVAGRL SHQK*

```

ORF62a and ORF62-1 show 98.9% identity in 284 aa overlap:

```

40     orf62a.pep  MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP  60
          |||||
          orf62-1  MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP  60

45     orf62a.pep  LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA 120
          |||||
          orf62-1  LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA 120

50     orf62a.pep  AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA 180
          |||||
          orf62-1  AAFAGVALLMAGGAEEGGEVGFGLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA 180

55     orf62a.pep  AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI 240
          |||||
          orf62-1  AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLCGWYAYWLWNKGMSRVPANVSGLLI 240

          orf62a.pep  SLEPVVGVLAVLILGEHLSPVSVLGVFVVIAATLVAGRLSHQKX  285
          |||||
          orf62-1  SLEPVVGVLAVLILGEHLSPVSVLGVFVVIAATLVAGRLSHQKX  285

```

60 Homology with a predicted ORF from *N.gonorrhoeae*

ORF62 shows 99.5% identity over a 216aa overlap with a predicted ORF (ORF62.ng) from *N. gonorrhoeae*:

	orf62.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGIKIPREEWKP	60
	orf62ng	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGIKIPREEWKP	60
5	orf62.pep	LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA	120
	orf62ng	LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA	120
10	orf62.pep	AAFAGVALLMAGGAEEGEVGWFGLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA	180
	orf62ng	AAFAGVALLMAGGAEEGEVGWFGLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA	180
	orf62.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCG	216
15	orf62ng	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI	240

The complete length ORF62ng nucleotide sequence <SEQ ID 247> is:

	1	ATGTTTACC	AAATCCTTGC	CCTGATTATC	TGGGGCAGCT	CGTTTATTGC
	51	CGCCAAATAT	GTCTATGGCG	GCATCGATCC	CGCATTGATG	GTCGGCGTGC
20	101	GCCTGCTGAT	TGCCGCGCTG	CCTGCACTGC	CCGCTGCGCG	CCGTCATGTC
	151	GGCAAGATTG	CGCGTGAGGA	ATGGAAGCCG	TTGCTGATTG	TGTCGTTTCTG
	201	CAACTATGTG	CTGACCCGTC	TGCTTCAGTT	TGTCGGGTTG	AAATACACTT
	251	CCGCGCGCCAG	CGCATCGGTC	ATTGTGCGAC	TCGAGCCGCT	GCTGATGGTG
	301	TTTGTGCGAC	ACTTTTCTT	CAACGACAAA	GCGCGTGCCT	ACCACTGGAT
	351	ATGCGGCGCG	GCGGCATTG	CCGGTGTCGC	GCTGCTGATG	GCGGGCGGTG
25	401	CGGAAGAGGG	CGGCGAAGTC	GGCTGGTTCG	GCTGCTGCT	GGTGTGTTG
	451	GCGGGCGCGG	GCTTTTGTG	CGCTATGCGT	CCGACGCAAA	GGCTGATTGC
	501	CCGCATCGGC	GCACCGGCAT	TCACATCTGT	TTCCATTGCC	GCCGCATCGT
	551	TGATGTGCTC	GCCGTTTTCG	CTTGCTTTGG	CGCAAAGTTA	TACCGTGGAC
	601	TGGAGCGTCG	GGATGGTATT	GTCGCTGTTG	TATTTGGGTT	TGGGGTGCGG
30	651	CTGGTACGCC	TATTGGCTGT	GGAACAAGGG	GATGAGCCGT	GTCCTGCCA
	701	ACGCGTCGGG	ACTGTTGATT	TCGCTCGAAC	CCGTGCTCGG	CGTGCTGTTG
	751	GCGGTTTTGA	TTTTGGGCGA	ACATTTATCG	CCCGTGTCG	CCTTGGGCGT
	801	GTTTGTCTG	ATCGCCGCCA	CTTTCGCCG	CGGCCGCTG	TCGCGCAGGG
	851	ACGCGCAAAA	CGGCAATGCC	GTCTGA		

35 This encodes a protein having amino acid sequence <SEQ ID 248>:

	1	MFYQILALII	WGSSFIAAKY	VYGGIDPALM	VGVRLLIAAL	PALPACRRHV
	51	GKIPREEWKP	LLIVSFVNYV	LTLLQLQFVL	KYTSAAASV	IVGLEPLLMV
	101	FVGHFFNDK	ARAYHWICGA	AAFAGVALLM	AGGAEEGEV	GWFGCLLVLL
40	151	AGAGFCAAMR	PTQRLIARIG	APFTSVSIA	AASLMCLPFS	LALAQSYTVD
	201	WSVGMVLSLL	YLGLGCGWYA	YWLWNKGMSR	VPANASGLLI	SLEPVVGVLL
	251	AVLILGEHLS	PVSALGVFVV	IAATFAAGRL	SRDAQNGNA	V*

ORF62ng and ORF62-1 show 97.9% identity in 283 aa overlap:

		10	20	30	40	50	60
45	orf62ng.pep	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGIKIPREEWKP					
	orf62-1	MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGIKIPREEWKP					
		10	20	30	40	50	60
50	orf62ng.pep	LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA					
	orf62-1	LLIVSFVNYVLTLLQLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFNDKARAYHWICGA					
		70	80	90	100	110	120
55	orf62ng.pep	AAFAGVALLMAGGAEEGEVGWFGLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA					
	orf62-1	AAFAGVALLMAGGAEEGEVGWFGLLVLLAGAGFCAAMRPTQRLIARIGAPFTSVSIA					
		130	140	150	160	170	180
60	orf62ng.pep	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI					
	orf62-1	AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANVSGLLI					
65		190	200	210	220	230	240

```

                250      260      270      280      290
orf62ng.pep    SLEPVVGVLLAVLILGEHLSPVSALGVFVVAATFAAGRLSRRDAQNGNAVX
5             orf62-1    SLEPVVGVLLAVLILGEHLSPVSALGVFVVAATLVAGRLSHQKX
                250      260      270      280

```

Furthermore, ORF62ng shows significant homology to a hypothetical *H.influenzae* protein:

```

10  sp|Q57147|Y976_HAEIN_HYPOTHETICAL_PROTEIN_HI0976 >gi|1074589|pir||B64163
    hypothetical protein HI0976 - Haemophilus influenzae (strain Rd KW20)
    >gi|1574004 (U32778) hypothetical [Haemophilus influenzae] Length = 128
    Score = 106 bits (262), Expect = 2e-22
    Identities = 56/114 (49%), Positives = 68/114 (59%)

15  Query: 1  MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRXXXXXXXXXXCRRHVGKIPREEWKP 60
        M YQILAL+IW SS I K Y +DP L+V VR R KI + K
    Sbjct: 1  MLYQILALLIWSSSLIVGKLTYSMDPVLVQVRLIAMIIVMPLFLRRWKKIDKPMRKQ 60

    Query: 61  LLIVSFVNYVLTLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDDKARAY 114
        L ++F NY LLQF+GLKYTSA+SA ++GLEPLL+VFVGHFFF K +
20  Sbjct: 61  LWWLAFFNYTAVFLLQFIGLKYTSAASAVTMIGLEPLLVVVFVGHFFFKTKQNGF 114

```

Based on this analysis, including the homology with the transmembrane protein of *H.influenzae* and the putative leader sequence and several transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 30

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 249>:

```

1  ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCmGwms TCCTGkkGTA
51  sGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
30 101  GGTGGATTGT TGCGTTCAGC GCAATGCTGC TGCTGGTGTT GTCGCGCGTT
    151  TTGGCAGCTT ATGTCAATATT GCTGTTGAAA GACAGGCGCG ACGGCGGTATT
    201  CGGTTGCGtA srTyGCCAAA gsGCCTgkks TGGG.ATGTT TACGCTGGTT
    251  GCGGCACTGC CCGGCGGTGT TCTGTTCCGC TTTCCCGCAC AGTTCATCAA
    301  CGGCACGATT AATTCGTGGT TCGGCAACGA TACCCACGAG GCGCTTGAAC
35 351  GCAGCCTCAA TTTGAGCAAG TCCGCATTGA ATTTGGCGGC AGACAACGCC
    401  CTCGCAACG CCGTCCCGT GCAGATAGAC CTCATCGGCG CGGCTTCCCT
    451  GCGGCGGAT ATGGGCGGG TGCTGGAACA TTACGCGGC AGCGGTTTTG
    501  CCCAGCTTGC CCTGTACAAy ksCGCAAGCG GCAAAATCGA AAAAAGCATC
    551  AACC CGCACA AGCTCGATCA GCGGTTTCCA GGTAAGGCGC GTTGGGAaAa
40 601  AATCCaACGG GCGGGTTCCG TCAGGGATTT GGAAAGCATA GCGGCGTAT
    651  TGTaCGCGCA GGGCTGGCTG TCGGCGGGTA CGCACwACGG GCGCGATTAC
    701  GCCTTGTTTT TCCGTACGCC GGTTCCTAAA GGCGTGCGAG AGGATGCCGT
    751  yTTAATCGAA AAGGCAAGGG CGAAATATGC TGAGTTGAGT TACAGCAAAA
    801  AAGGTTTGCA GACCTTTTTC CTGGCAACCC TGCTGATTGC CTCGCTGCTG
45 851  TCGATTTTTC TTGCACTGGT CATGGCACTG TATTTCCGCC GCGGTTTCGT
    901  CGAACCCTGC CTATCGCTTG CCGAGGGGGC GAAGGCGGTG GCGCAAGGCG
    951  ATTTACGCCA GACGCGCCCC GTGTTGCGCA ACGACGAGTT CGGACGCTTG
1001 ACCArGTTGT TCAACCACAT GACCGAGCAG CTTTCCATCG CCAAAGATGC
1051 AGACGAGCGC AACCGCCGGC GCGAGGAAGC CGCCAGGCAT TATCTTGAAT
50 1101 GCGTGTGGA GGGGCTGACC ACGGGCGTGG TGGTGTTTGA CGAACAAGGC
    1151 TGTCTGAAA CCTTCAACAA AGCGGCGGGT ACC..

```

This corresponds to the amino acid sequence <SEQ ID 250; ORF64>:

```

1  MRRFLPIAAI CAXXLXXGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
55 51  LARYVILLK DRRDGVFGSX XAKXPXXMF TLVAXLPGVF LFGFPAQFIN
    101  GTINSWFGND THEALERSLN LSKSALNLAA DNALGNAVVP QIDLIGAASL
    151  PGDMGRVLEH YAGSGFAQLA LYNXASGKIE KSINPHKLDQ PFPKARWEK
    201  IQRAGSVRDL ESIGGVLYAQ GWLSAGTHXG RDYALFFRQP VPKGVAEDAV
    251  LIEKARAKYA ELSYSKGLQ TFFLATLLIA SLLSIFLALV MALYFARRFV

```

301 EPVLSLAEGA KAVAQGDFSQ TRPVLNRNDEF GRLTXLFNHM TEQLSIAKDA
 351 DERNRRREEA ARHYLECVLE GLTTGVVVFDE EQGCLKTFNK AAGT..

Further work revealed the complete nucleotide sequence <SEQ ID 251>:

1 ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCCGTCG TCCTGTTGTA
 5 51 CGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
 101 GGTGGATTGT TGCATTACG GCAATGCTGC TGCTGGTGTG GTCCGCCGTT
 151 TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACAGCGTATT
 201 CGGTTTCGAG ATTGCCAAAC GCCTTTCTGG GATGTTTACG CTGGTTGCCG
 251 TACTGCCCGG CGTGTTCCTG TTCGGCGTTT CCGCACAGTT CATCAACGGC
 10 301 ACGATTAATT CGTGGTTCGG CAACGATACC CACGAGGCGC TTGAACGCAG
 351 CCTCAATTG AGCAAGTCCG CATTGAATTT GCGGCGAGAC AACGCCCTCG
 401 GCAACGCCGT CCCCCTGCAG ATAGACCTCA TCGGCGCGGC TTCCTTGCCC
 451 GGGGATATGG GCAGGGTGCT GGAACATTAC GCCGCGAGCG GTTTTGCCCA
 501 GCTTGCCCTG TACAATGCCG CAAGCGGCAA AATCGAAAA AGCATCAACC
 15 551 CGCACAAAGCT CGATCAGCGG TTCCAGGTA AGGCGCGTTG GGAAAAATC
 601 CAACGGGCGG GTTCGGTCAG GGATTGGAA AGCATAGGCG GCGTATTGTA
 651 CGCGCAGGCG TGCTGTTCGG CGGGTACGCA CAACGGGCGC GATTACGCTT
 701 TGTTTTTCCG TCAGCCGGTT CCCAAAGGCG TGGCAGAGGA TGCCGCTCTA
 751 ATCGAAAAGG CAAGGGCGAA ATATGCTGAG TTGAGTTACA GCAAAAAAGG
 20 801 TTGTCAGACC TTTTCTCTGG CAACCTGCTG GATTGCTCG CTGCTGTGCA
 851 TTTTCTTTC ACTGCTCATG GCACTGTATT TCGCCGCGCG TTTGCTCGAA
 901 CCCGTCTTAT CGCTTGCCGA GGGGCGAAG GCGGTGGCGC AAGGCGATT
 951 CAGCCAGACG CGCCCGCTGT TGCACAACGA CGAGTTCGGA CGCTTGACCA
 1001 AGTTGTTCAA CCACATGACC GAGCAGCTT CCATCGCCAA AGAAGCAGAC
 25 1051 GAGCGCAACC GCCGCGCGGA GGAAGCCGCC AGGCATTATC TTGAATGCGT
 1101 GTTGGAGGGG CTGACCACGG GCGTGGTGGT GTTGACGAA CAAGGCTGTC
 1151 TGAAAACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC
 1201 CCCCTGTGGG GCAGCAGCGG GCACGTTGG CACGCGCTTT CGGCGCAGCA
 1251 GTCCCTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGCGC GCAGGTACGG
 30 1301 ACAAACCGGT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG
 1351 CTGGGCAAGG CAACCGTCCT GCCGAAGAC AACGGCAACG GCGTGGTAAT
 1401 GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAA GAAGCCGCGT
 1451 GGGGCGAAGT GGCGAAGCGG CTGGCACACG AAATCCGAA TCCGCTCACG
 1501 CCCATCCAGC TTTCCGCCGA ACGGCTGGCG TGGAATTTGG GCGGGAAGCT
 35 1551 GGATGAGCAG GATGCGCAAA TCCTGACGCG TTCGACCGAC ACCATCGTCA
 1601 AACAGGTGGC GGCATTGAAG GAAATGGTCG AAGCATTCCG CAATTATGCG
 1651 CGTTCCCTT CGCTCAAATT GGAATTCAG GATTTGAACG CCTTAATCGG
 1701 CGATGTGTTG GCATTGTATG AAGCCGGTCC GTGCCGGTTT GCGGCGGAGC
 1751 TTGCCGGCGA ACCGCTGACG GTGGCGGCGG ATACGACCGC CATGCGGCAG
 40 1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
 1851 TGTGCCCGAA GTCAGGGTAA AATCGGAAAC AGGGCAGGAC GGTGCGATTG
 1901 TCCTGACGGT TTGCGACAAC GGCAAAGGT TCGGCAGGGA AATGCTGCAC
 1951 AACGCCTTCG AGCCGTATGT AACGGACAAA CCGGCGGGAA CGGGATTGGG
 2001 TCTGCCTGTG GTGAAAAAAA TCATTGAAGA ACACGGCGGC CGCATCAGCC
 45 2051 TGAGCAATCA GGATGCGGGT GCGCGGTGTG TCAGAATCAT CTTGCCAAAA
 2101 ACGGTAAAAA CTTATGCGTA G

This corresponds to the amino acid sequence <SEQ ID 252; ORF64-1>:

1 MRRELPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
 51 LARYVILLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFING
 101 TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAPVQ IDLIGAASLP
 151 GDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQ FPGKARWEKI
 201 QRAGSVRDLE SIGGVLYAQ WLSAGTHNGR DYALFFRQPV PKGVAEDAVL
 251 IEKARAKYAE LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARRFVE
 301 PVLSLAEGAK AVAQGDFSQT RPVLNRNDEF RLTKLFNHMT EQLSIAKEAD
 351 ERNRRREEA RHYLECVLEG LTTGVVVFDE QGCLKTFNKA AEQILGMPLT
 401 PLWGSRRHWG HGVSAAQSL AEFVFAAGAA AGTDKPVHVK YAAPDDAKIL
 451 LGKATVLPED NGNGVMVID DITVLIHAQ EAAWGEVAKR LAHEIRNPLT
 501 PIQLSAERLA WKLGGKLEQ DAQILTRSTD TIVKQVAALK EMVEAFRNYA
 551 RSPSLKLENQ DLNALIGDVL ALYEAGPCRF AAELAGEPLT VAADTTAMRQ
 601 VLHNIFKNAA EAAEEADVPE VRVKSETGQD GRIVLTVCDN GKGFREMLH
 651 NAFEPVYTDK PAGTGLGLPV VKKIIEHGG RISLSNQDAG GACVRILPK
 701 TVKTYA*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF64 shows 92.6% identity over a 392aa overlap with an ORF (ORF64a) from strain A of *N.*

meningitidis:

5	orf64.pep	10 20 30 40 50 60	MRRFLPIAAICAXLXXGLTAATGSTSSSLADYFWWIVAFSAML	LLVLSAVLARYVILLK
	orf64a	10 20 30 40 50 60	MRRFLPIAAICAVVLLYGLTAATGSTSSSLADYFWWIVAFSAML	LLVLSAVLARYVILLK
10	orf64.pep	70 80 90 100 110 120	DRRDGVFGSXXAKXPXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN	
	orf64a	70 80 90 100 110	DRRDGVFGSQIAKR-LSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLN	
15	orf64.pep	130 140 150 160 170 180	LSKSALNLAADNALGNAVFPVQIDLIGAASLPDGMGRVLEHYAGSGFAQLALYNXASGKIE	
	orf64a	120 130 140 150 160 170	LSKSALNLAADNALGNAIPVQIDXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIE	
20	orf64.pep	190 200 210 220 230 240	KSINPHKLDQFFPGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP	
	orf64a	180 190 200 210 220 230	KSINPHKLDQFFPGKARWEKIQAGSVRDLESIGGVLYAXGWLSAXTHNGRDYALFFRQP	
30	orf64.pep	250 260 270 280 290 300	VPKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV	
	orf64a	240 250 260 270 280 290	VPKGVAEDAVLIEKARAXXXLSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV	
35	orf64.pep	310 320 330 340 350 360	EPVLSLAEGAKAVAQGDQSOTRPLRNDEFGRLTXLFNHMTQLSIAKDADERNRRREEA	
	orf64a	300 310 320 330 340 350	EPVLSLAEGAKAVAQGDQSOTRPLRNDEFGRLTXLFNHMTQLSIAKEADERNRRREEA	
40	orf64.pep	370 380 390	ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAGT	
	orf64a	360 370 380 390 400 410	ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAEQILGMPLTPLWGSSRHGWHGVSAQQSL	
45	orf64a	420 430 440 450 460 470	LAEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNXNGVVMVDDITVLIHAQ	

The complete length ORF64a nucleotide sequence <SEQ ID 253> is:

50	1	ATGCGCCGTT	TTCTACCGAT	CGCAGCCATA	TGCGCCGTCG	TCCTGTTGTA
	51	CGGACTGACG	GCGGCAACCG	GCAGCACCAG	TTCGCTGGCG	GATTATTCTT
	101	GGTGGATTGT	TGCGTTCAGC	GCAATGCTGC	TGCTGGTGTT	GTCCGCCGTT
	151	TTGGCACGTT	ATGTCATATT	GCTGTTGAAA	GACAGGCGCG	ACGGCGTATT
	201	CGGTTCCGAG	ATTGCCAAAC	GCCTTTCGGG	GATGTTTACG	CTGGTTGCCG
55	251	TACTGCCCGG	CGTGTTTCTG	TTCGGCGTTT	CCGCACAGTT	TATCAACGGC
	301	ACGATTAATT	CGTGGTTCGG	CAACGATACC	CACGAGGCGC	TTGAACGCAG
	351	CCTCAATTG	AGCAAGTCCG	CATTGAATCT	GGCGGCAGAC	AACGCCCTTG
	401	GCAACGCCAT	CCCCGTGCAG	ATAGACNTCA	TCGGCGCGGC	TTCCCTGCC
	451	NGGGATATGG	GCAGGGTGCT	GGAACATTAC	GCCGGCAGCG	GTTTTGCCCA
60	501	GCTTGCCCTG	TACAATGCCG	CAAGCGGCAA	AATCGAAAAA	AGCATCAACC
	551	CGCACAAAGC	CGATCAGCCG	TTTCCAGGTA	AGCGCGTTG	GGAAAAATC
	601	CAACAGGCGG	GTTCCGTCAG	GGATNNGGAA	AGCATAGGCG	GCGTATTGTA
	651	CGCGCANGGC	TGGCTGTCGG	CAGNNACGCA	CAACGGGCGC	GATTACGCCT
	701	TGTTTTTCCG	TCAGCCGTT	CCCAAAGGCG	TGGCAGAGGA	TGCCGTCTTA
65	751	ATCGAAAAGG	CAAGGGCGNA	ANANNNTNAG	TTGAGTTACA	GCAAAAAAGG
	801	TTTGCAGACC	TTTTCTCTNG	CAACCCTGCT	GATTGCCTCN	CTGCTGTGCA
	851	TTTTCTTGC	ACTGGTCATG	GCACTGTATT	TCGCCGCCCG	TTTCGTGCGA

5
 10
 15
 20
 25

```

901 CCGGTCCTAT CGCTTGCCGA GGGGCGAAG GCGGTGGCGC AAGGCGATT
951 CAGCCAGACG CGCCCCGTGT TGCACACGA CGAGTTCCGA CGCTTGACCA
1001 AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
1051 GAGCGCAACC GCCGGCGCGA GGAAGCCGCC AGACATTATC TCGAATGCGT
1101 GTTGGAGGGG CTGACCACGG GCGTGGTGGT GTTTGACGAA CAAGGCTGTC
1151 TGAAAACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC
1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
1251 GTCCCTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGCGC GCAGGTACGG
1301 ACAAACCGGT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG
1351 CTGGGCAAGG CAACCGTCTT GCCCGAAGAC AACNGCAACG GCGTGGTAAT
1401 GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAA GAAGCCGCGT
1451 GGGGCGAAGT GGCAAAACGG CTGGCACACG AAATCCGCAA TCCGCTCACG
1501 CCCATCCAGC TTTCTGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
1551 GGACGAGCAN GACGCGCAAA TCCTGACACG TTCGACCGAC ACCATCATCA
1601 AACAAGTGGC GGCATTAAAA GAAATGGTCG AGGCATTCCG CAATTACNCG
1651 CGTTCCCTT CGNCTCAATT GGAAATCAG GATTGAACG CCTTAATCGG
1701 CGATGTGTTG GCATTGTACG AAGCTGGTCC GTGCCGGTTT GCGGCGGAAC
1751 TTGCCGCGCA ACCGCTGATG ATGGCGCGG ATACGACCGC CATGCGGCAG
1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGCGG AAGAAGCCGA
1851 TGTGCCCGAA GTCAGGGTAA AATCGGAAGC GGGGCGAGAC GGACGGATTG
1901 TCCTGACAGT TTGCGACAAC GGCAAGGGGT TCGGCGAGGA AATGCTGCAC
1951 AATGCCTTCG AGCCGTATGT AACGGACAAA CCGGCTGGAA CGGGATTGNG
2001 ACTGCCCGTG GTGAAAAAAA TCATTGAAGA ACACGGCGGC CNCATCAGCC
2051 TGAGCAATCA GGATGCGGGC GCGCGTNTG TCAGAATCAT CTTGCCAAAA
2101 ACGGTAGAAA CTTATGCGTA G
  
```

This encodes a protein having amino acid sequence <SEQ ID 254>:

30
 35
 40

```

1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
51 LARYVILLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFING
101 TINSWFGND HEALERSLNL SKSALNLAAD NALGNAIPVQ IDXIGAASLP
151 XDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI
201 QQAGSVRDXE SIGGVLYXG WLSAXTHNGR DYALFFRQPV PKGVAEDAVL
251 IEKARAXXXX LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARRFVE
301 PVLSLAEGAK AVAQGDFSQT RPLVRNDEFG RLTKLFNHMT EQLSIAKEAD
351 ERNRRREEAA RHYLECVLEG LTTGVVVFDE QGCLKTFNKA AEQILGMPLT
401 PLWGSSRHGW HGVSAQQSLL AEFVAAIGAA AGTDKPVHVK YAAPDDAKIL
451 LGKATVLPED NXNGVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT
501 PIQLSAERLA WKLGGKLDX DAQILTRSD TTIKQVAALK EMVEAFRNYX
551 RSPSXQLENQ DLNALIGDVL ALYEAGPCRF AAELAGEPIM MAADTTAMRQ
601 VLHNIKNAE EAAEADVPE VRVKSEAGQD GRIVLTVCDN GKGFGREMLH
651 NAFEPYVTDK PAGTGLXLPV VKKIEEHGG XISLSNQDAG GAXVRIILPK
701 TVETYA*
  
```

ORF64a and ORF64-1 show 96.6% identity in 706 aa overlap:

45
 50
 55
 60
 65

```

              10      20      30      40      50      60
orf64a.pep  MRRFLPIAAICAVVLLYGLTAATGSTSSLA DYFWWIVAFS AMLLLVLSAVLARYVILLK
orf64-1      MRRFLPIAAICAVVLLYGLTAATGSTSSLA DYFWWIVAFS AMLLLVLSAVLARYVILLK
              10      20      30      40      50      60

              70      80      90      100     110     120
orf64a.pep  DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
orf64-1      DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
              70      80      90      100     110     120

              130     140     150     160     170     180
orf64a.pep  SKSALNLAADNALGNAIPVQIDXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIEK
orf64-1      SKSALNLAADNALGNAIPVQIDLIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIEK
              130     140     150     160     170     180

              190     200     210     220     230     240
orf64a.pep  SINPHKLDQFPFGKARWEKIQQAGSVRDXE SIGGVLYXGWLSAXTHNGRDYALFFRQPV
orf64-1      SINPHKLDQFPFGKARWEKIQQAGSVRDLESIGGVLYXGWLSAXTHNGRDYALFFRQPV
              190     200     210     220     230     240

              250     260     270     280     290     300
  
```

	orf64a.pep	PKGVAEDAVLIEKARAXXXLSYSKKGLOTFFLATLLIASLLSIFLALVMALYFARRFE
	orf64-1	PKGVAEDAVLIEKARAKYAELSYSKKGLOTFFLATLLIASLLSIFLALVMALYFARRFE
5		250 260 270 280 290 300
	orf64a.pep	310 320 330 340 350 360
	orf64-1	PVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRITKLFNHMTQLSIAKEADERNRRREEAA
10		310 320 330 340 350 360
	orf64a.pep	370 380 390 400 410 420
	orf64-1	RHYLECVLEGLTTGVVVFDEQGCKLTFNKAAEQILGMPLTPLWGSSRHGWHGVSAAQSSLL
15		370 380 390 400 410 420
	orf64a.pep	430 440 450 460 470 480
	orf64-1	AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNKGVMVIDDITVLIHAQK
20		430 440 450 460 470 480
	orf64a.pep	490 500 510 520 530 540
	orf64-1	EAANGVAKRLAHEIRNPLTPIQLSAERLAWKLGKLDQDAQILTRSTDITIKQVAALK
25		490 500 510 520 530 540
	orf64a.pep	550 560 570 580 590 600
	orf64-1	EMVEAFRNYXRSPSKQLENQDLNALIGDVLALYEAGPCRFAAELAGEPLMMAADTTAMRQ
30		550 560 570 580 590 600
	orf64a.pep	610 620 630 640 650 660
	orf64-1	VLHNI FKNAAEAAEEADVPEVRVKSEAGQDGRIVLTVCDNGKGF GREMLHNAFEPYVTDK
35		610 620 630 640 650 660
	orf64a.pep	670 680 690 700
	orf64-1	PAGTGLXLPVVKKIIIEHGGXISLSNQDAGGAXVRIILPKTVETYAX
40		670 680 690 700
	orf64a.pep	PAGTGLGLPVVKKIIIEHGGRIISLSNQDAGGACVRIILPKTVKTYAX
45		670 680 690 700

Homology with a predicted ORF from *N.gonorrhoeae*

ORF64 shows 86.6% identity over a 387aa overlap with a predicted ORF (ORF64.ng) from *N.*

50 *gonorrhoeae*:

	orf64.pep	MRRFLPIAIAICAXLXXGLTAATGSTSSLADYFWWIVAFSAML LLLVLSAVLARYVILLK	60
	orf64ng	MRRFLPIAIAICAVVLLYGLTAATGSTSSLADYFWWIVSFSAML LLLVLSAVLARYVILLK	60
55	orf64.pep	DRRDGVFGSXXAKXPXXXMFTLVAXLPGVFLGFPQAFINGTINSWFGNDTHEALERSLN	120
	orf64ng	DRRNGVFGSQIAKR-LSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLN	119
60	orf64.pep	LSKSALNLAADNALGNAPVQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNXASGKIE	180
	orf64ng	LSKSALDLAADNAVSNAPVQIDLIGTASLSGNMGVLEHYAGSGFAQLALYNASGKIE	179
	orf64.pep	KSINPHKLDQPPGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP	240
65	orf64ng	KSINPHQFDQPLPDKEHWEIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYALFFRQP	239

-185-

	orf64.pep	VPKGAEDAVLIEKARAKYAELSYSKGLQTFFLATLLIASLLSIFLALVMALYFARRFV	300
	orf64ng	IPENVAQDAVLIEKARAKYAELSYSKGLQTFFLVTLIASLLSIFLALVMALYFARRFV	299
5	orf64.pep	EPVLSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTXLFNHMTQLSIAKDADERNRRREEA	360
	orf64ng	EPILSLAEGAKAVAQGDFSQTRPVLNRNDEFGRLTCLFNHMTQLSIAKEADERNRRREEA	359
10	orf64.pep	ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAGT	394
	orf64ng	ARHYLECVLDGLTTGVVVSYP LSCCRTAVFSTCHSSPLSYF	400

An ORF64ng nucleotide sequence <SEQ ID 255> was predicted to encode a protein having amino acid sequence <SEQ ID 256>:

15	1	MRRFLPIAAI	CAVVLLYGLT	AATGSTSSLA	DYFWWIVSFS	AMLLVL SAV
	51	LARYVILLK	DRRNGVFGSQ	IAKRLSGMFT	LVAVLPGFL	FGISAQFING
	101	TINSWFGNDT	HEALERSLNL	SKSALDLAAD	NAVSNAPVQ	IDLIGTASLS
	151	GNMGSVLEHY	AGSGFAQLAL	YNAASGKIEK	SINPHQFDQP	LPDKEHWEQI
20	201	QQTGSVRSLE	SIGGVLYAQG	WLSAGTHNGR	DYALFFRQPI	PENVAQDAVL
	251	IEKARAKYAE	LSYSKGLQTF	FFLVTLIAS	LLSIFLALVM	ALYFARRFVE
	301	PILSLAEGAK	AVAQGDFSQT	RPVLRNDEF	RLTKLFNHMT	EQLSIAKEAD
	351	ERNRRREEAA	RHYLECVLDG	LTTGVVVSYP	LSCCRTAVFS	TCHSSPLSYF*

Further work revealed the complete gonococcal DNA sequence <SEQ ID 257>:

25	1	ATGCGCGCT	TCCTACCGAT	CGCAGCCATA	TGCGCGCTCG	TCCTGCTGTA
	51	CGGATTGACG	GCGGCGACCG	GCAGCACCAG	TTCGCTGGCG	GATTATTTCT
	101	GGTGGATAGT	CTCGTTACAG	GCAATGCTGC	TGCTGGTGTT	GTCCGCGGTT
	151	TTGGCACGTT	ATGTCATATT	GCTGTTGAAA	GACAGGCGCA	ACGGCGTGTT
	201	CGGTTTCGAG	ATTGCCAAAC	GCCTTTCCGG	GATGTTACAG	CTGGTCGCGG
	251	TACTGCCCGG	CTTGTTCTCT	TTCGGCATT	CCGCGCAGTT	TATCAACGGC
30	301	ACGATTAATT	CGTGGTTCGG	CAACGACACC	CACGAAGCCC	TGAACGCAG
	351	CCTTAATTG	AGCAAGTCCG	CACTGGATT	GCGGCGAGAC	AATGCCGTCA
	401	GCAACGCCGT	TCCCGTACAG	ATAGACCTCA	TCGGCACCAG	CTCCCTGTCT
	451	GGCAATATGG	GCAGTGTGCT	GGAACACTAC	GCCGGCAGCG	GTTTTGCCCA
	501	GCTTGCCCTG	TACAATGCCG	CAAGCGGGAA	AATCGAAAAA	AGCATCAATC
	551	CGCACCAATT	CGACGACCG	CTTCCCGACA	AAGAACATTG	GGAACAGATT
35	601	CAGCAGACCG	GTTCGGTTCG	GAGTTTGGA	AGCATAGGCG	GCGTATTGTA
	651	CGCGCAGGGA	TGTTGTCTCG	CAGGTACGCA	CAACGGGCGC	GATTACGCGC
	701	TGTTCTTCCG	CCAGCCGATT	CCCGAAAAATG	TGGCACAGGA	TGCCGTTCTG
	751	ATTGAAAAGG	CGCGGGCGAA	ATATGCCGAA	TTGAGTTACA	GCAAAAAGG
40	801	TTTGACAGAC	TTTTTTCTGG	TAACCTGTCT	GATTGCCTCG	CTGCTGTCTGA
	851	TTTTTCTTGC	GCTGGTAATG	GCACTGTATT	TTGCCGCGCG	TTTCGTCTGAA
	901	CCCATTCTGT	CGCTTGCCGA	GGGCGCAAAG	GCGGTGGCGC	AGGGTGATT
	951	CAGCCAGACG	GCCTCCGTAT	TGCGCAACGA	CGAGTTCGGA	CGTTTGACCA
	1001	AGCTGTTCAA	CCATATGACC	GAGCAGCTTT	CCATCGCCAA	AGAAGCAGAC
	1051	GAACGCAACC	GCCGCGCGCA	GGAAGCCGCG	CGTCACTACC	TCGAGTGCGT
45	1101	GTTGGATGGG	TTGACTACCG	GTGTGGTGGT	GTTTGACGAA	AAAGGCCGTT
	1151	TGAAAACCTT	CAACAAGCGG	GCGGAACAGA	TTTTGGGGAT	GCCGCTCGCC
	1201	CCCCTGTGGG	GCAGCAGCCG	GCACGGTTGG	CACGGCGTTT	CGGCGCAGCA
	1251	GTCCCTGCTT	GCCGAAGTGT	TtgcgcgcAT	CGGTGCGGCG	GCAGGTACGG
50	1301	ACAAACCGGT	CCAGGTGGAA	TATGCCGCGC	CGGACGATGC	CAAAATCCTG
	1351	CTGGGCAAGG	CGACGGTATT	GCCCGAAGAC	AACGGCAACG	GCGTGGTGAT
	1401	GGTGATTGAC	GACATACCGG	TGCTGATACG	CGCGCAAAAA	GAAGCCGCGT
	1451	GGGGTGAAGT	GGCGAAGCGG	CTGGCACACG	AAATCCGCAA	TCCGCTCAGC
	1501	CCCATCCAGC	TTTCCGCCGA	ACGGCTGGCG	TGGAATTGG	GCGGGAAGCT
	1551	GGACGATCAG	GACGCGCAAA	TCCTGACGCG	TtcgACCGAC	ACCATCATCA
55	1601	AACAGgtggc	gGCGTTAAAA	GAAATGGTCG	AGGCATTCCG	CAATTACGCG
	1651	CGCGCCCTT	CGCTCAAAT	GGAATATCAG	GATTTGAACG	CCTTAATCGG
	1701	CGATGTTTTG	GCCCTGTACG	AAGCCGGCCC	GTGCCGGTTT	GAGGCGGAAC
	1751	TTGCCGGCGA	ACCGCTGATG	ATGGCGGCGG	ATACGACCGC	CATGCGGCAG
60	1801	GTGCTGCACA	ATATTTTCAA	AAATGCCGCG	GAAGCGGCGG	AAGAAGCCGA
	1851	TATGCCCGAA	GTCAGGTA	AATCGGAAAC	GGGGCGAGGAC	GGACGGATTG
	1901	TCCTGACGGT	TTGCCGACAAC	GGCAAGGGAT	TCGGCAAGGA	AATGCTGCAC
	1951	AATGCTTTCG	AGCCGTATGT	GACGGATAAG	CCGGCGGGAA	CGGGACTGGG
	2001	TCTGCCTGTA	GTGAAAAAAA	TCATTGGAGA	ACACGGCGGC	CGCATCAGCC
	2051	TGAGCAATCA	GGATGCGGGT	GGGGCGGTG	TCAGAATCAT	CTTGCCAAAA
65	2101	ACGGTAGAAA	CTTATGCGTA	G		

This corresponds to the amino acid sequence <SEQ ID 258; ORF64ng-1>:

```

1  MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVSFS AMLLLVLSAV
51  LARYVILLLK DRRNGVFGSQ IAKRLSGMFT LVAVLPGLFL FGISAQFING
101 TINSWFGNDT HEALERSLNL SKSALDLAAD NAVSNAVVPQ IDLIGTASLS
151 GNMGSVLEHY AGSGFAQLAL YNAASGKIEK SINPHQFDQP LPDKEHWEQI
201 QQTGSVRSLE SIGGVLYAQQ WLSAGTHNGR DYALFFRQPI PENVAQDAVL
251 IEKARAKYAE LSYSKKGLQT FFLVTLLIAS LLSIFLALVM ALYFARRFVE
301 PILSLAEGAK AVAQGDFSQT RPVLRNDEFG RLTKLFNHMT EQLSIAKEAD
351 ERNRRREEAA RHYLECVLDG LTTGVVVFDE KGRKTFNKA AEQILGMPLA
401 PLWGSSRHGW HGVSAQOSLL AEVFAAIGAA AGTDKPVQVE YAAPDDAKIL
451 LGKATVLPED NGNGVVMVID DITVLIRAQK EAAWGEVAKR LAHEIRNPLT
501 PIQLSAERLA WKLGGKLDQ DAQILTRSTD TTIKQVAALK EMVEAFRNYA
551 RAPSLKLENQ DLNALIGDVL ALYEAGPCRF EAELAGEPLM MAADTTAMRQ
601 VLHNIKFNAE EAAEEADMPE VRVKSETGQD GRIVLTVCDN GKGFGKEMLH
651 NAFEPYVTDK PAGTGLGLPV VKKIIGEHHG RISLSNQDAG GACVRIILPK
701 TVETYA*

```

ORF64ng-1 and ORF64-1 show 93.8% identity in 706 aa overlap:

```

20 orf64ng-1.pep 10 20 30 40 50 60
    MRRFLPIAAICAVVLLYGLTAATGSTSSLA DYFWWIVSFSAMLLLVLSAVLARYVILLLK
    orf64-1      10 20 30 40 50 60
    MRRFLPIAAICAVVLLYGLTAATGSTSSLA DYFWWIVAFSAMLLLVLSAVLARYVILLLK

25 orf64ng-1.pep 70 80 90 100 110 120
    DRRNGVFGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNL
    orf64-1      70 80 90 100 110 120
    DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL

30 orf64ng-1.pep 130 140 150 160 170 180
    SKSALDLAADNAVSNVAVPQIDLIGTASLSGNMGSVLEHYAGSGFAQLALYNAASGKIEK
    orf64-1      130 140 150 160 170 180
    SKSALNLAADNALGNVAVPQIDLIGAASLPQDMGRVLEHYAGSGFAQLALYNAASGKIEK

35 orf64ng-1.pep 190 200 210 220 230 240
    SINPHQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQQWLSAGTHNGRDYALFFRQPI
    orf64-1      190 200 210 220 230 240
    SINPHKLDQPPFGKARWEKIQRAGSVRDLESIGGVLYAQQWLSAGTHNGRDYALFFRQPV

40 orf64ng-1.pep 250 260 270 280 290 300
    PENVAQDAVLIEKARAKYAE LSYSKKGLQTFFLVTLLIASLLSIFLALVMALYFARRFVE
    orf64-1      250 260 270 280 290 300
    PKGVAEDAVLIEKARAKYAE LSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE

45 orf64ng-1.pep 310 320 330 340 350 360
    PILSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMT EQLSIAKEADERNRRREEAA
    orf64-1      310 320 330 340 350 360
    PVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMT EQLSIAKEADERNRRREEAA

50 orf64ng-1.pep 370 380 390 400 410 420
    RHYLECVLDGLTTGVVVFDEKGRKTFNKA AEQILGMPLAPLWGSSRHGW HGVSAQOSLL
    orf64-1      370 380 390 400 410 420
    RHYLECVLEGLTTGVVVFDEQGLKTFNKA AEQILGMPLT PLWGSSRHGW HGVSAQOSLL

55 orf64ng-1.pep 430 440 450 460 470 480
    AEVFAAIGAAAGTDKPVQVEYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIRAQK
    orf64-1      430 440 450 460 470 480
    AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIHAQK

60 orf64ng-1.pep 490 500 510 520 530 540
    EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDQDAQILTRSTD TTIKQVAALK
    orf64-1      490 500 510 520 530 540
    EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDQDAQILTRSTD TTIKQVAALK

```

5	orf64-1	EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKGLDEQDAQILTRSTDITIVKQVAALK	490	500	510	520	530	540
	orf64ng-1.pep	EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAEELAGEPLMMAADTTAMRQ	550	560	570	580	590	600
10	orf64-1	EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFAAEELAGEPLTVAADTTAMRQ	550	560	570	580	590	600
	orf64ng-1.pep	VLHNIFKNAAEAAEADMPEVRVKSETGQDGRIVLTVCDNGKGFGEMLHNAFEPYVTDK	610	620	630	640	650	660
15	orf64-1	VLHNIFKNAAEAAEADMPEVRVKSETGQDGRIVLTVCDNGKGFGEMLHNAFEPYVTDK	610	620	630	640	650	660
	orf64ng-1.pep	PAGTGLGLPVVKKIIEHGGRISLSNQDAGGACVRIILPKTVETYAX	670	680	690	700		
20	orf64-1	PAGTGLGLPVVKKIIEHGGRISLSNQDAGGACVRIILPKTVKTYAX	670	680	690	700		

Furthermore, ORF64ng-1 shows significant homology to a protein from *A. caulinodans*:

25	sp Q04850 NTRY_AZOCA NITROGEN REGULATION PROTEIN NTRY >gi 77479 pir S18624 ntry protein - Azorhizobium caulinodans >gi 38737 (X63841) NtrY gene product [Azorhizobium caulinodans] Length = 771 Score = 218 bits (550), Expect = 7e-56 Identities = 195/720 (27%), Positives = 320/720 (44%), Gaps = 58/720 (8%)							
	Query: 7	IAAICAVVLLYGLTAATGSTSSLDYFWIXXXXXXXXXXXXXXXXXRYVILLKDRNGV	66					
30		I+A+ ++L GLT + + + R + + K R G						
	Sbjct: 35	ISALATFLILMGLTPVPTHQVVIS----VLLVNAAVLILSAMVGREIWRIAKARAGR	90					
35	Query: 67	FGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNLSKSALD	126					
	Sbjct: 91	AAARLHIRIVGLFAVSVVPAILVAVVASLTLDRLDRWFSMRTQEIIVASSVSVAQTYVR	150					
40	Query: 127	LAADNAVSNAPVQIDLIGTASLSGNMGSVLEHYAG--SGFAQLALYNAASGKIEKSINP	184					
	Sbjct: 151	EHALNIRGDILAMSADLTRLKSV-----YEGDRSRFNQILTAQAALRNLPGLAMLI	200					
45	Query: 185	HQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYA-----	233					
	Sbjct: 201	RR-DLSVVERAN-VNIGREFIVPANLAIGDATPDQFVIYLP--NDADYVAAVVPLKDYDD	256					
50	Query: 234	--LFFRQPIPENVAQDAVLIEKARAKYAEISYSKKGLQTFFLVTVXXXXXXXXXXXXXVMA	291					
	Sbjct: 257	LYLYVARLIDPRVIGYLKTTQETLADYRSLEERRFGVQVAFALMYAVITLIVLLSAVWLG	316					
55	Query: 292	LYFARRFVEPILSLAEGAKAVAQGFDSQTRPVLRND-EFGRLTKLFNHMTQELSIXXXXX	350					
	Sbjct: 317	LNFSKWLVAPIRRLMSAADHVAEGLDVRVPIYRAEGDLASLAETFNKMTHELRSQREAI	376					
60	Query: 351	XXXXXXXXXXXXHYLECVLDGLTTGVVVFDEKGRKLTFNKAAEQILGMPLAPLWGSSRHGW	410					
	Sbjct: 377	LTARDQIDSRRRFTEAVLSGVGAGVIGLDSQERITILNRSERLLG--LSEVEALHRHLA	434					
65	Query: 411	HGVSAQSSLLAEVFXXXXXXXXTDKPVQVEYAAPDDAKILLGKATVLPEDNG---NGVVM	467					
	Sbjct: 435	EVVPETAGLLEEA-----EHARQRSVQGNITLTRDGRERVFVAVRVTEQSPAEHGWVV	488					
70	Query: 468	VIDDITVLIRAQKEAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGKGLDDQDAQILTR	527					
	Sbjct: 489	TLDDITELISAQRTSAWADVARRIAHEIKNPLTPIQLSAERLKRKFGRHV-TQDREIFDQ	547					
75	Query: 528	STDITIKQVAALKEMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAEELAGE	587					
	Sbjct: 548	CTDTIIRQVGDIGRMVDEFSSFARMKPVVDSQDMSEIIRQTVFLMRVGHPEVDFSEVP	607					
80	Query: 588	PLMMAA-DTTAMRQVLHNIFKNXXXXXXXXDMPEVRVK-----SETGQDGRIVLTVCD	639					
	Sbjct: 608	PAMPARFDRRLVSQALTNILKNAAEAEI-EAVP-PDVRGQGRIRVSANRVGED--LVIDIID	664					

Query: 640 NGKGFGEKMLHNAFEPYVTDKPGTGLGLPVVKIIGEHHGRISLSNQDAG-GACVRIIL 698
 NG G +E + EPYVT + GTGLGL +V KI+ EHGG I L++ G GA +R+ L
 Sbjct: 665 NGTGLPQESNRRLLEPYVTTREKGTGLGLAIVGKIMEEHGGGIELNDAPEGRGAWIRLTL 724

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 31

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 259>:

```

1 ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
51 GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101 CTTTCCAAAT TTTCCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC
151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCC GCCCTT TTGCTTTCCT
251 ACGTCTTTTC CGTTTGTTC CACAACGGCA GTTGGACAGG CTTGGGCGCG
301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC
351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTGTATT CACAAATTAC
401 GCCGTCTGAA AGCGTGTTGG ATTGCACCGA ACGCATCAAC CGTCATCGGG
451 CACGCGTTGG ATACG...

```

This corresponds to the amino acid sequence <SEQ ID 260; ORF66>:

```

1 MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTWGAFSFP
51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRLKAWW IAPNASTVIG
151 HALDT...

```

Further work revealed the complete nucleotide sequence <SEQ ID 261>:

```

1 ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
51 GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101 CTTTCCAAAT TTTCCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC
151 TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
201 GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCC GCCCTT TTGCTTTCCT
251 ACGTCTTTTC CGTTTGTTC CACAACGGCA GTTGGACAGG CTTGGGCGCG
301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC
351 CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTGTATT CACAAATTAC
401 GCCGTCTGAA AGCGTGTTGG ATTGCACCGA CCGCATCAAC CGTCATCGGC
451 AACGCCTTGG ATACGCTGGT ATTTTTCGCC GTTGCCTTCT ACGCAAGCAG
501 CGATGGATT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
551 TGTTCAAAC TACCGTCTGC ACCCTCTTCT TCCTGCCCGC CTACGGCGTG
601 ATACTGAATC TGCTGACGAA AAACTGACA ACCCTGCAAA CCAAACAGGG
651 GCAAGACCGC CCCGCGCCCT CGCTGCAAAA TCCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 262; ORF66-1>:

```

1 MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTWGAFSFP
51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRLKAWW IAPTASTVIG
151 NALDTLVFFA VAFYASSDGF MAANWQGI AFVDYLFKLTVC TLFFLPAYGV
201 ILNLLTKKLT TLQTKAQDR PAPSLQNP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o221 of *E. coli* (accession number P37619)

50 ORF66 and o221 protein show 67% aa identity in 155aa overlap:

```

orf66 1 MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV 60
      M F+ Q+ KALF L LFH+L+I +SNYLQ P I G HTTWGAFSFPFIFLATDLTV
o221 1 MNVFSQTQRYKALFWLSLFHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV 60

5 orf66 61 RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
      RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
o221 61 RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFALAHFNLFVARIATASFMAYA 120

10 orf66 121 IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT 155
      +GQILD+ VFN+LR+ + WW+AP AST+ G+ DT
o221 121 LGQILDVHVFNRLRQSRWWLAPTASTLFGNVSDT 155

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF66 shows 96.1% identity over a 155aa overlap with an ORF (ORF66a) from strain A of *N.*

15 *meningitidis*:

```

              10      20      30      40      50      60
orf66.pep  MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV
              |||
orf66a     MYAFTAAQQQKALFWLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV
              10      20      30      40      50      60

              70      80      90      100     110     120
orf66.pep  RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA
              |||
orf66a     RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA
              70      80      90      100     110     120

              130     140     150
orf66.pep  IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT
              :|||
orf66a     LGQILDIFVFNKLRLKAWWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIAP
              130     140     150     160     170     180

orf66a     VDYLEFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX
              190     200     210     220

```

The complete length ORF66a nucleotide sequence <SEQ ID 263> is:

```

1  ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCTGGCTGGT
51  GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101 CCTTCCAAAT TTCCGGCATC CACACCACTT GGGGCGCGTT TTCCTTTCCC
140 151 TTCATCTTCC TCGCCACCGA CCTGACCGTC CGCATTTTCG GTTCGCACTT
201 GGCACGCGCG ATTATCTTTT GGGTCATGTT CCCCGCCCTT TTGCTTTCCT
251 ACGTCTTTTC CGTTTGTTC CACAACGGCA GTTGACGGG CTGGGCGCG
301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCGCTGG CAAGTTTTCG
45 351 CGCCTACGCG CTCGGACAAA TCCTTGATAT TTTTGTGTTT AACAAATTAC
401 GCGTCTGAA AGCGTGGTGG GTTGCCCCGA CTGCATCAAC CGTCATCGGC
451 AACGCCTTAG ATACGTTGGT ATTTTTCGCC GTTGCCTTCT ACGCAAGCAG
501 CGATGGATT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
551 TGTTCAAAC CACCGTCTGC GGTCTGTTT TCCTGCCCGC CTACGGCGTG
601 ATTCTGAATC TGCTGACGAA AAACTGACG ACCCTGCAAA CCAAACAGGC
50 651 GCAAGACCGC CCCGCGCCCT CGCTGCAAAA TCCGTAA

```

This encodes a protein having amino acid sequence <SEQ ID 264>:

```

1  MYAFTAAQQQ KALFWLVLFH ILIIAASNYL VQFPFQISGI HTTWGAFSFP
51  FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA LGQILDIFVF NKLRLKAWW VAPTASTVIG
151 NALDTLVFFA VAFYASSDGF MAANWQGIAP VDYLEFKLTV CGLFFLPAYGV
201 ILNLLTKKLT TLQTKQAQDR PAPSLQNP*

```

ORF66a and ORF66-1 show 97.8% identity in 228 aa overlap:

```

              10      20      30      40      50      60
orf66a.pep MYAFTAAQQQKALFWLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV
              |||
orf66-1     MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV

```

-190-

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf66a.pep	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA					
	orf66-1	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA					
		70	80	90	100	110	120
10	orf66a.pep	LGQILDIFVFNKLRLKAWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF					
	orf66-1	IGQILDIFVFNKLRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF					
		130	140	150	160	170	180
15	orf66a.pep	VDYLFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX					
	orf66-1	VDYLFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX					
		190	200	210	220	229	
20							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF66 shows 94.2% identity over a 155aa overlap with a predicted ORF (ORF66.ng) from *N.*

gonorrhoeae:

25	orf66.pep	MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV	60
	orf66ng	MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV	60
	orf66.pep	RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA	120
30	orf66ng	RIFGSHLARRIIFWVMFPALSLSYVFSVLFHNGSWTGLGAPSQFNTFVGRIALASFAAYA	120
	orf66.pep	IGQILDIFVFNKLRLKAWWIAPNASTVIGHALDT	155
	orf66ng	LGQILDIFVFDKLRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF	180

35 The complete length ORF66ng nucleotide sequence <SEQ ID 265> is:

1	ATGTACGCAT	TGACCGCCGC	ACAGCAACAG	AAGGCACTCT	TCCGGCTGGT
51	GCTTTTCCAT	ATCCTCATCA	TCGCCGCCAG	CAACTATCTG	GTGCAGTTCC
101	CCTTCCGGAT	TTTCGGCATC	CACACCACTT	GGGGCGCGTT	TTCCTTTCCC
151	TTCATCTTCC	TCGCCACCGA	CCTGACCGTC	CGCATTTCG	GTTCCGACTT
201	GGCGCGCGCG	ATTATCTTTT	GGGTGATGTT	CCCCGCCCTT	ttgCTTcat
251	aCGTCTTTC	CGTTTGTTC	CACAACGGCA	GTGGACGGG	CTGGGCGCG
301	ctgTCCCAAT	TCAACACCTT	TGTCGGACGC	ATCGCGCTGG	CAAGTTTTC
351	CGCCTACGCG	CTCGGACAAA	TCCTTGATAT	TTTCGTATTC	GACAAATTAC
401	GCCGTCTGAA	AGCGTGGTGG	ATTGCCCGCG	CCGCATCAAC	CGTCATCGGC
451	AATGCACTGG	ACACGTIAGT	ATTTTTTGCC	GTTCCTTTT	ACGCAAGCAG
501	CGATGAATTT	ATGGCGGCAA	ACTGGCAGGG	CATCGCTTTT	GTTCGATTACC
551	TGTTCAAAC	TACCGTCTGC	ACCCTCTTCT	TCCTGCCCGC	CTACGGCGTG
601	ATACTGAATC	TGCTGACGAA	AAACTGACG	GCCCTGCAAA	CCAAACAGGC
651	GCAAGACCGC	CCCGTGCCCT	CGCTGCAAAA	TCCGTAA	

50 This encodes a protein having amino acid sequence <SEQ ID 266>:

1	MYALTAAQQQ	KALFRLVLFH	ILIIAASNYL	VQFPFRIFGI	HTTWGAFSFP
51	FIFLATDLTV	RIFGSHLARR	IIFWVMFPAL	LSYVFSVLF	HNGSWTGLGA
101	PSQNTFVGR	IALASFAAYA	LGQILDIFVF	DKLRLKAWW	IAPAASTVIG
151	NALDTLVFFA	VAFYASSDEF	MAANWQGI AF	VDYLFKLTVC	TLFFLPAYGV
201	ILNLLTKKLT	ALQTKQAQDR	PVPSLQNP*		

An alternative annotated sequence is:

1	MYALTAAQQQ	KALFRLVLFH	ILIIAASNYL	VQFPFRIFGI	HTTWGAFSFP
51	FIFLATDLTV	RIFGSHLARR	IIFWVMFPAL	LSYVFSVLF	HNGSWTGLGA
101	LSQNTFVGR	IALASFAAYA	LGQILDIFVF	DKLRLKAWW	IAPAASTVIG
151	NALDTLVFFA	VAFYASSDEF	MAANWQGI AF	VDYLFKLTVC	TLFFLPAYGV
201	ILNLLTKKLT	ALQTKQAQDR	PVPSLQNP*		

ORF66ng and ORF66-1 show 96.1% identity in 228 aa overlap:

```

5  orf66-1.pep  MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV 60
    orf66ng    MYALTAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV 60
    orf66-1.pep  RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
    orf66ng    RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120
10  orf66-1.pep  IGQILDIFVFNKLRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGI AF 180
    orf66ng    LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF 180
15  orf66-1.pep  VDYLFLKLTVC TLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX 229
    orf66ng    VDYLFLKLTVC TLFFLPAYGVILNLLTKKLTALQTKQAQDRPVPSLQNPX 229

```

Furthermore, ORF66ng shows significant homology with an *E.coli* ORF:

```

20  spIP37619|YHHQ_ECOLI HYPOTHETICAL 25.3 KD PROTEIN IN FTSY-NIKA INTERGENIC
    REGION (O221)
    >gi|1073495|pir||S47690 hypothetical protein o221 - Escherichia coli >gi|466607
    (U00039) No definition line found [Escherichia coli] >gi|1789882 (AE000423)
    hypothetical 25.3 kd protein in ftsY-nika intergenic region [Escherichia coli]
    Length = 221
    Score = 273 bits (692), Expect = 5e-73
25  Identities = 132/203 (65%), Positives = 155/203 (76%)

    Query: 1 MYALTAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV 60
        M + Q+ KALF L LFH+L+I +SNYLQV P I G HTTWGAFSFPFIFLATDLTV
    Sbjct: 1 MNVFSQTRQYKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV 60
30  Query: 61 RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120
        RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
    Sbjct: 61 RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFALAHFNLFVARIATASF MAYA 120
35  Query: 121 LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGI AF 180
        LGQILD+ VF++LR+ + WW+AP AST+ GN DTL FF +AF+ S D FMA +W IA
    Sbjct: 121 LGQILDVHVFNRLRQSRWWLAPTASTLFGNVSDTLAFFFIAFWRSPDAFMAEHWMEIAL 180
40  Query: 181 VDYLFLKLTVC TLFFLPAYGVILN 203
        VDY FK+ + +FFLP YGV+LN
    Sbjct: 181 VDYCFKVLISIVFFLPMYGVLLN 203

```

Based on this analysis, including the homology with the *E.coli* protein and the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 32

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 267>:

```

50  1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
    51 AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAAyGCA GTmwrAATAT
    101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
    151 GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
    201 TTTAACACAC AyyCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
    251 CCGCCAGCGT ATCCCGCGCC GCGTATTGG CGGGGGTCGG CAACTTGCC
55  301 CGCTTAGCG CGAAATTCAG CACAAGGCG GTtCCCTATG TCGGAACAGC
    351 CcTTTAGGCC CACGACGTAT ACGAAAcTTT CAAAGAAGAC ATACAGGCAC
    401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTGTAAA AGGCTACGAA
    451 TATAGTAATT GCCTTTGGTA CGAAGACAAA AGACGTATTA ATAGAACCTA

```

501 TGGCTGCTAC GGC GTT GAT..

This corresponds to the amino acid sequence <SEQ ID 268; ORF72>:

1 MVIKYTNLNF AKLSIIAILM MYSFEANANA VXISETVSVD TGQGAKIHKF
51 VPKNSKTYSS DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGVKLA
5 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFKVGYE
151 YSNCLWYEDK RRINRTYGCY GVD..

Further work revealed the complete nucleotide sequence <SEQ ID 269>:

1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
51 AATTTTGATG ATGTATTTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
10 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
151 GTACCTAAAA ATAGTAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
201 TTTAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGG CGGGGGTCTGG CAAACTTGCC
301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC
15 CTTTTAGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC
401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTGCAAA GGTCTCAGGC
451 TAA

This corresponds to the amino acid sequence <SEQ ID 270; ORF72-1>:

1 MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF
20 51 VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVKLA
101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVS
151 *

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

25 ORF72 shows 98.0% identity over a 147aa overlap with an ORF (ORF72a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf72.pep		MVIKYTNLNF AKLSIIAILM MYSFEANANA VXISETVSVD TGQGAKIHKF VPKNSKTYSS					
30 orf72a		MVIKYTNLNF AKLSIIAILM MYSFEANANA VXISETVSVD TGQGAKIHKF VPKNSKTYSS					
		10	20	30	40	50	60
		70	80	90	100	110	120
orf72.pep		DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGKLARL GAKFSTRAPV PYVGTALLA					
35 orf72a		DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGKLARL GAKFSTRAPV PYVGTALLA					
		70	80	90	100	110	120
		130	140	150	160	170	
40 orf72.pep		HDVYETFKEDI QARGYQYDP ETDKFKVGYE YSNCLWYEDK RRINRTYGCY GVD					
orf72a		HDVYETFKEDI QARGYQYDP ETDKFAKVS GX					
		130	140	150			

The complete length ORF72a nucleotide sequence <SEQ ID 271> is:

45 1 ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
51 AATTTTGATG ATGTATTTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
101 CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
151 GTACCTAAAA ATAGTAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
201 TTTAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
50 251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGG CGGGGGTCTGG CAAACTTGCC
301 CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC
351 CTTTTAGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC
401 GAGGCTACCA ATACGACCCC GAAACCGACA AATTGCAAA GGTCTCAGGC
451 TAA

55 This encodes a protein having amino acid sequence <SEQ ID 272>:

```

1  MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF
51  VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVKLA
101 RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVS
151 *

```

5 ORF72a and ORF72-1 show 100.0% identity in 150 aa overlap:

```

10 orf72a.pep      10      20      30      40      50      60
    MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF VPKNSKTYSS
    |||
10 orf72-1        10      20      30      40      50      60
    MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF VPKNSKTYSS
    |||

15 orf72a.pep      70      80      90      100     110     120
    DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKFSTRA VPYVGTALLA
    |||
15 orf72-1        70      80      90      100     110     120
    DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKFSTRA VPYVGTALLA
    |||

20 orf72a.pep      130     140     150
    HDVYETFKED IQARGYQYDP ETDKFAKVS GX
    |||
20 orf72-1        130     140     150
    HDVYETFKED IQARGYQYDP ETDKFAKVS GX
    |||

```

Homology with a predicted ORF from *N.gonorrhoeae*

25 ORF72 shows 89% identity over a 173aa overlap with a predicted ORF (ORF72.ng) from *N. gonorrhoeae*:

```

30 orf72.pep      MVIKYTNLNF AKLSIIAILM MYSFEANANA VXiSETVSVD TGQGAKIHKF VPKNSKTYSS 60
    || :|||
30 orf72ng        MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAKVHKF VPKSSNIYSS 60
    || :|||

35 orf72.pep      DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAVGVK LARLGAKFSTRA VPYVGTALLA 120
    || :|||
35 orf72ng        DLTkAVDLTH IPTGAKARIN AKITASVSRA GVLAVGVK LVRQGAkfGTRAVPYVGTALLA 120
    || :|||

40 orf72.pep      HDVYETFKED IQARGYQYDP ETDKfVKGYEYSNCLWYEDKRRINRTYGCYGV D 173
    |||
40 orf72ng        HDVYETFKED IQARGCRYDP ETDKfVKGYEYANCLWYEDERRINRTYGCYGV DSSIMRLM 180
    |||

```

An ORF72ng nucleotide sequence <SEQ ID 273> was predicted to encode a protein having amino acid sequence <SEQ ID 274>:

```

40 1  MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAKVHKF
    51  VPKSSNIYSS DLTkAVDLTH IPTGAKARIN AKITASVSRA GVLAVGVKLV
    101 RQGAkFGTRA VPYVGTALLA HDVYETFKED IQARGCRYDP ETDKfVKGYE
    151 YANCLWYEDE RRINRTYGCY GVDSSIMRLM PDRSRfPEVK QLMESQMYRL
    201 ARPFWNWRKE ELNKLSSLDW NNFVLNRCTF DWNGGCAVN KGDDFRAGAS
    45 251 FSLGRNPkYK EEMDAKPEE ILSLKV DADP DKYIEATGYF GYSEKVEVAP
    301 GTKVNMGPVT DRNGNPVQVA ATfGRDAQGN TTADVQVIPR PDLT PASAEA
    351 PHAQPLPEVS PAENPANNDP PDENPGTRPN PEPD PDLNPD ANPD TDGQPG
    401 TSPDSPAVPD RPNGRHRKER KEGEDGGLSC DYfPEILACQ EMGKPSDRMF
    451 HDISIPQVTD DKTWSSHNF LPSNGVCPQPK TFHVfGRQYR ASYEP LCVFA
    50 501 EKIRFAVLLA FIIMSAfVVF GSLGGE*

```

After further analysis, the following gonococcal DNA sequence <SEQ ID 275> was identified:

```

55 1  ATGGTCACAA AACATACAAA TTGAATTTT GCGAAATTGT CGATAATTGC
    51  AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAATGCA GTAAAAATAT
    101 CTGAAACTCT TTCGGTTGAT ACCGACAAG GCGCGAAAGT TCATAAGTTC
    151 GTTCCTAAAT CAAGTAATAT TTATT CATCT GATTTAACAA AAGCGGTAGA
    201 TTTAACGCAT ATCCCCACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
    251 CCGCCAGCGT ATCCCGCGCC GCGGTATTGT CGGGGGTCGG CAAACTTGTC
    301 CGCCAAGGCG CGAAATTCGG CACAAGGCG GTTCCCTATG TCGGAACAGC
    351 CCTTTTAGCC CACGACGTAT ACGAACTTT CAAAGAAGAC ATACAGGCAC
    60 401 GAGGCTGCCG ATACGATCCC GAAACGACA AATT

```


5 ORF72ng-1 and ORF721-1 show 89.7% identity in 145 aa overlap:

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 277>:

35 This corresponds to the amino acid sequence <SEQ ID 278; ORF73>:

Further work revealed the complete nucleotide sequence <SEQ ID 279>:

This corresponds to the amino acid sequence <SEQ ID 280; ORF73-1>:

BNSDOCID: <WO 8924578A2.1_>

151 SRNAIEHKKD E*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF73 shows 90.8% identity over a 76aa overlap with an ORF (ORF73a) from strain A of *N.*

5 *meningitidis:*

```

10      10      20      30      40      50      60
orf73.pep  MRFFGIGFLVLLFLEIMSVVWADWLGGGWTFLMAAGFAAGVLMRLQTGLTGLLLAGAA
          |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
orf73a     MRFFGIGFLVLLFLEIMSVVWADWLGGGWTFLMAATFAAGVVMRLRHTGLSGLLAGAA
          |||||:||||:||||:||||:||||:||||:||||:||||:||||:
15      70
orf73.pep  MRSGGKVSVYQMLWPI
          |||||:||||:||||:||||:||||:||||:||||:||||:||||:
orf73a     MRSGGRVSVYXMLWXIRYTVAAVCXMSPGFVSSVXAVLLXLPFKGGAVLQAGGAENFFNM
          |||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

The complete length ORF73a nucleotide sequence <SEQ ID 281> is:

	1	ATGAGATTTT	TCGGTATCGG	TTTTTTGGTG	CTGCTGTTTT	TGGAGATTAT
	51	GTCGATTGTG	TGGGTTGCCG	ATTGGTTGGG	CGCGCGTTGG	ACGCTGTTTC
20	101	TAATTGGCGC	AACCTTTGCC	GCCGGCGTGG	TGATGCTCAG	GCATACGGTT
	151	CTGTCCGGTC	TTTTATTGGC	GGGCGCGGCA	ATGAGAAAGC	GCGGGAGGGT
	201	ATCCGTTTAT	CANATGTTGT	GGCNTATCCG	TTATACGGTG	GCGGCGGTGT
	251	GTCTGATGAG	TCCGGGATTC	GTATCCTCGG	TGCTNGCCGT	ATTGCTGNTG
	301	CTNCCGTTTA	AGGGAGGTGC	AGTGTTCGAG	CAGGAGAGTG	CGGAAAATTT
25	351	TTTCAACATG	AACCCNTCGG	GCAGAAAAGA	NGGCNTTCC	CGCGATGACG
	401	ATATTATCGA	GGGGGAATAT	ACGGTTGAAG	ANCCTTACGG	CGGCANTCGT
	451	TTCCGAAACG	CCNTNGAACA	CAAAAAAGAC	GAATAA	

This encodes a protein having amino acid sequence <SEQ ID 282>:

30

1	MRFFGIGFLV	LLFLEMSIV	WVADWLGGGW	TLFLMAATFA	AGVVMLRHTG
51	LSGLLLAGAA	MRSGRVSVY	XMLWXIRYTV	AAVCXMSGPF	VSSVXAVLLX
101	LPFKGGAVLQ	AGGAENFFNM	NXSGRKXGXS	RDDDIIEGEY	TVEXPYGGXR
151	FRNAXEHKKD	E*			

ORF73a and ORF73-1 show 91.3% identity in 161 aa overlap

		10	20	30	40	50	60
35	orf73a.pep	MRFFGIGFLVLLFLEIMSVVWADWLGGGWTFLFMAATFAAGVVMRLRHTGLSGLLLAGAA					
	orf73-1	MRFFGIGFLVLLFLEIMSVVWADWLGGGWTFLFMAAGFAAGVLMRLRHTGLSGLLLAGAA					
		10	20	30	40	50	60
		70	80	90	100	110	120
40	orf73a.pep	MRSGGRVSVYXMLWXIRYTVAAVCXMSPGFVSSVXAVLLXLPFKGGAVLQAGGAENFFNM					
	orf73-1	MRSGGRVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFKGGAVLQAGGAENFFNM					
		70	80	90	100	110	120
45		130	140	150	160		
	orf73a.pep	NXSGRKXGXSRRDDIIIEGYTVEXPYGGXRFNRNAXEHKKDEX					
	orf73-1	NQSGRKEGFSRDDIIIEGYTVEEPYGGNRSRNAIEHKKDEX					
		130	140	150	160		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF73 shows 92.1% identity over a 76aa overlap with a predicted ORF (ORF73.ng) from *N.*

gonorrhoeae:

55 orf73.pep MRFFGIGFLVLLFLEIMSIWVADWLGGGWTLFLMAAGFAAGVLMRLQTGLTGLLLAGAA 60
|||||:|:|:|

```

orf73ng      MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAATFAAGVLMRLHTGLSGLLLAGAA 60
orf73.pep    MRSGBKSVYQMLWPI 76
5 orf73ng    VKSSGKSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFKGGAVLQAGGAENFFNM 120

```

The complete length ORF73ng nucleotide sequence <SEQ ID 283> is:

```

1 ATGAGATTTT TCGGTATCGG TTTTGTGGTG CTGCTGTTTT TGGAAATTAT
51 GTCGATTGTG TGGGTGCGG ATTGGCTGGG CGGCGGTTGG AcgcTGTTTC
101 TAATGGCGGC AACCTTTGCC GCCGGTGTGC TGATGCTCAG GCATAcggGG
151 CTGTCCGGTC TTTATTGGC TGGCGCGGCG GTAAAAagta gtgGGAAGGT
201 ATCTGTTTAT CagatgtTGT GGCCTATCCG TTATAcggtg gcggcggtgt
251 GTCTGatgag tCcgGATTC GTATCCTccg tgttggCGGT ATTGCTGCTG
301 CTGCCgttta aggGaggGgc agtgttgca gaggagggtg cggaaaATTT
351 TTTCAACATg aaCcaatcgg gcagaaAaga gggatttttc cacgatgacg
15 401 atattatcga gggagaatat acggttgaaa aacctgacgg cggcaatcgt
451 tcccgaAAcg ccatcgaaca cgaaaAgac gaataA

```

This encodes a protein having amino acid sequence <SEQ ID 284>:

```

1 MRFFGIGFLV LLFLEIMSIW VVADWLGGGW TLFLMAATFA AGVLMRLHTG
51 LSGLLLAGAA VKSSGKSVY QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
20 101 LPFKGGAVLQ AGGAENFFNM NQSGRKEGFF HDDDIIEGEY TVEKPDGGNR
151 SRNAIEHEKD E*

```

ORF73ng and ORG73-1 show 93.8% identity in 161 aa overlap

```

25 orf73-1.pep MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAAGFAAGVLMRLHTGLSGLLLAGAA
orf73ng MRFFGIGFLVLLFLEIMSIWVADWLGGGWTFLFLMAATFAAGVLMRLHTGLSGLLLAGAA
10 20 30 40 50 60
30 orf73-1.pep MRSGBRVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFKGGAVLQAGGAENFFNM
orf73ng VKSSGKSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFKGGAVLQAGGAENFFNM
70 80 90 100 110 120
35 orf73-1.pep NQSGRKEGFSRDDDIIEGEYTVPEPYGGNRSRNAIEHKKDEX
orf73ng NQSGRKEGFFHDDDIIEGEYTVKPDGGNRSRNAIEHEKDEX
130 140 150 160

```

40 Based on this analysis, including the presence of a putative leader sequence and putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 34

45 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 285>:

```

1 ATGTTTGT TTTCAGACGGC ATTCTT.ATG TTTCAGAAAC ATTTGCAGAA
51 AGCCTCCGAC AGCGTCGTCG GAGGGACATT ATACGTGGTT GCCACGCCCA
101 TCGGCAATTT GCGGACATT ACCCTGCGCG CTTTGGCGGT ATTGCAAAAG
151 GCG..... GCGA AGACACGCGC GTTACCGCAC AGCTTTTGAG
50 201 CGCGTACGGC ATTCAGGGCA AACTCGTCAG TGTGCGCGAA CACAACGAAC
251 GGCAGATGGC GGACAAGATT GTCGGCTATC TTTCAGACGG CATGGTTGTG
301 GCACAGGTTT CCGATGCGGG TACGCCGGCC GTGTGCGACC CGGGCGCGAA
351 ACTCGCCCGC CGCGTGCCTG AGGCCGGGTT TAAAGTCGTT CCCGTCGTGG
401 GCGCAAC.GC GGTGATGGCG GCTTTGAGCG TGGCCGGTGT GGAAGGATCC
55 451 GATTTTATT TCAACGGTTT TGTACCGCCG AAATCGGGAG AACGCAGGAA

```

-197-

501 ACTGTTTGCC AAATGGGTGC GGGCGGCGTT TCCTATCGTC ATGTTTGAAA
 551 CGCCGCACCG CATCGGTGCA GCGCTTGCCG ATATGGCGGA ACTGTTCCCC
 601 GAACGCCGAT TAATGCTGGC GCGCGAAATT ACGAAAACGT TTGAAACGTT
 651 CTTAAGCGGC ACGGTTGGGG AAATTCAGAC GGCATTGTCT GCCGACGGCG
 701 ACCAATCGCG CGGCGAGATG GTGTTGGTGC TTTATCCGGC GCAGGATGAA
 751 AAACACGAAG GCTTGTCGCA GTCCGCGCAA AACATCATGA AAATCCTCAC
 801 AGCCGAGCTG CCGACCAAAC AGGCGGCGGA GCTTGCTGCC AAAATCACGG
 851 GCGAGGGAAA GAAAGCTTTG TACGAT..

This corresponds to the amino acid sequence <SEQ ID 286; ORF75>:

10 1 MFVFQTAFXM FQKHLQKASD SVVGGTLYV V ATPIGNLADI TLRALAVLQK
 51 A...AEDTR VTAQLLSAYG IQGKLVSVRE HNERQMADKI VGYLSDGMVV
 101 AQVSDAGTFA VCDPGAKLAR RVREAGFKV PVVGAXAVMA ALSVAGVEGS
 151 DFYFNGFVFP KSGERRKLF KVVRAAFPIV MFETPHRIGA ALADMAELFP
 201 ERRMLLAREI TKTFTFLSG TVGEIQTALS ADGDQSRGEM VLVLYPAQDE
 15 251 KHEGLSESAQ NIMKILTAEL PTKQAELAA KITGEGKKAL YD..

Further work revealed the complete nucleotide sequence <SEQ ID 287>:

1 ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
 51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCTGCG
 101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG
 151 CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATTACAGG GCAAACCTCGT
 201 CAGTGTGCGC GAACACAACG AACGGCAGAT GCGGACAAAG ATTGTCGGCT
 251 ATCTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG
 301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG
 351 GTTTAAAGTC GTTCCCCTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA
 401 GCGTGGCCCG TGTGGAAGGA TCCGATTTT ATTCAACGG TTTGTACCG
 451 CCGAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TCGGGGCGGC
 501 GTTTCCTATC GTCATGTTT AAACGCCGCA CCGCATCGGT GCGACGCTTG
 551 CCGATATGGC GGAAGTGTTC CCCGAACGCC GATTAATGCT GCGCGCGGAA
 601 ATTACGAAAA CGTTTGAAC GTTCTTAAGC GGCACGCTTG GGGAAATTCA
 651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTTGG
 701 TGCTTTATCC GCGCGAGGAT GAAAAACAG AAGGCTTGTC CGAGTCCGCG
 751 CAAAACATCA TGAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCCGC
 801 GGAGCTTGCT GCCAAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC
 851 TGGCTCTGTC TTGAAAAAC AAATAG

35 This corresponds to the amino acid sequence <SEQ ID 288; ORF75-1>:

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
 51 RVTAQLLSAY GIQKLVSVR EHNERQMA DKIVGYLSDGMV VAQVSDAGTP
 101 AVCDPGAKLA RRVREAGFKV VPVGASAVM AALSVAGVEG SDFYFNGFVP
 151 PKSGERRKLF AKVVRAAFPI VMFETPHRIG ATLDMAELF PERRMLLARE
 201 ITKTFTFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA
 251 QNIMKILTAE LPTKQAELA AKITGEGKKA LYDLALSWKN K*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF75 shows 95.8% identity over a 283aa overlap with an ORF (ORF75a) from strain A of *N.*

45 *meningitidis*:

		10	20	30	40	50	60
orf75.pep		MFVFQTAFXM	FQKHLQKASD	SVVGGTLYV	VATPIGNLAD	ITLRALAVLQ	KAXXXXAEDTR
orf75a							
50			MFQKHLQKASD	SVVGGTLYV	VATPIGNLAD	ITLRALAVLQ	KADIICAEDTR
			10	20	30	40	50
orf75.pep		70	80	90	100	110	120
		VTAQLLSAYGI	IQGKLVSVRE	HNERQMA DKIV	GYLSDGMV	VAQVSDAGT	PAVCDPGAKLAR
55							
orf75a		VTAQLLSAYGI	IQGKLVSVRE	HNERQMA DKIV	GYLSDGMV	VAQVSDAGT	PAVCDPGAKLAR
		60	70	80	90	100	110
		130	140	150	160	170	180
orf75.pep		RVREAGFKV	VPVGAXAVMA	AALSVAGVEG	SDFYFNGFV	PPKSGERRKL	FAKVVRAAFPIV

The complete length ORF75a nucleotide sequence <SEQ ID 289> is:

This encodes a protein having amino acid sequence <SEQ ID 290>:

ORF75a and ORF75-1 show 98.3% identity in 291 aa overlap:

BNSDOCID: <WO__9924578A2_1_>

	orf75-1	ATLADMAELFPERRRLMLAREITKTFETFLSGTVGEIQTALSADGNQSRGEMVLVLYPAQD					
		190	200	210	220	230	240
5	orf75a.pep	250 260 270 280 290					
		EKHEGLSESAQNIMKILTAE LPTKQAAELAAKITGEGKKALYDLALSWKNKX					
	orf75-1	EKHEGLSESAQNIMKILTAE LPTKQAAELAAKITGEGKKALYDLALSWKNKX					
		250	260	270	280	290	

10 Homology with a predicted ORF from *N.gonorrhoeae*

ORF75 shows 93.2% identity over a 292aa overlap with a predicted ORF (ORF75.ng) from *N.*

gonorrhoeae:

15	orf75.pep	MFVFQTAFFM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQK-----AEDTR	56
	orf75ng	MSVFQTAFFM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQKADIICAEDTR	60
20	orf75.pep	VTAQLLSAYGIQGLVSVREHNERQMA DKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR	116
	orf75ng	VTAQLLSAYGIQGLVSVREHNERQMA DKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLAR	120
25	orf75.pep	RVREAGFKVVPVVGAXAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLFAKQVRAAFPV	176
	orf75ng	RVREAGFKVVPVVGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFAKQVRAAFPV	180
30	orf75.pep	MFETPHRIGAAALADMAELFPERRRLMLAREITKTFETFLSGTVGEIQTALSADGQSRGEM	236
	orf75ng	MFETPHRIGATLADMAELFPERRRLMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEM	240
35	orf75.pep	VLVLYPAQDEKHEGLSESAQNIMKILTAE LPTKQAAELAAKITGEGKKALYD	288
	orf75ng	VLVLYPAQDEKHEGLSESAQNAMKILAELPTKQAAELAAKITGEGKKALYDLALSWKNK	300

An ORF75ng nucleotide sequence <SEQ ID 291> was predicted to encode a protein having amino acid sequence <SEQ ID 292>:

35	1	MSVFQTAFFM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQK
	51	ADIIICAEDTR VTAQLLSAYG IQGLVSVRE HNERQMA DKV IGFLSDGLV
40	101	AQVSDAGTPA VCDPGAKLAR RVREAGFKVV PVVGASAVMA ALSVAGVAES
	151	DFYFNGFVPP KSGERRKLFA KQVRAAFPVV MFETPHRIGA TLADMAELFP
	201	ERRRLMLAREI TKTFTFLSG TVGEIQTALA ADGNQSRGEM VLVLYPAQDE
	251	KHEGLSESAQ NAMKILAEL PTKQAAELAA KITGEGKKAL YDLALSWKNK
	301	*

After further analysis, the following gonococcal DNA sequence <SEQ ID 293> was identified:

45	1	ATGTTTCAGA AACACTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
	51	ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCAGAC ATTACCCCTGC
50	101	GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATTGTGTC CGAAGACACG
	151	CGCGTTACTG CGCAGCTTTT GAGCGCGTAC GGCATTACAG GCAGGTGGGT
	201	CAGTGTGCGC GAACACAACG AGCGGCAGAT GCGGACAAAG GTAATCGGTT
	251	TCCTTTTCAGA CGGCCTGGTT GTGGCGCAGG TTTCCGATGC GGGTACGCCG
55	301	GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GCGAAGCAGG
	351	GTTCAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTAATG GCGGCGTTGA
	401	GTGTGGCCGG TGTGGCGGAA TCCGATTTT ATTTCACCG TTTTGTACCG
	451	CCGAAATCGG GCGAACGTAG GAAATTGTTT GCCAAATGGG TGCGGGCGCG
60	501	ATTTCCCTGTC GTCATGTTTG AAACGCCGCA CCGAATCGGG GCAACGCTTG
	551	CCGATATGGC GGAATTGTTT CCCGAACGCC GTCTGATGCT GCGCGCGGAA
	601	ATCACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTC
	651	GACGGCATTG GCGGCGGACG GCAACCAATC GCGCGGCGAG ATGGTGTGG
65	701	TGCTTTATCC GCGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCTGCG
	751	CAAAATGCGA TGAAATCCTT TGCGGCCGAG CTGCCGACCA AGCAGCGCGC
	801	GGAGCTTGCC GCCAAGATTA CAGGTGAGGG CAAAAGGCT TTGTACGATT
	851	TGGCACTGTC GTGGA AAAAC AAATGA

60 This corresponds to the amino acid sequence <SEQ ID 294; ORF75ng-1>:

-200-

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITRLALAVLQ KADIICAEDT
 51 RVTAQLLSAY GIQGLVSVR EHNERQMA DK VIGFLSDGLV VAQVSDAGTP
 101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP
 151 PKSGERRKLF AKWVRAAFV VMFETPHRIG ATLADMAELF PERRMLLARE
 5 201 ITKTFFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
 251 QNAMKILAAE LPKQAAELA AKITGEGKKA LYDLALSWKN K*

ORF75ng-1 and ORF75-1 show 96.2% identity in 291 aa overlap:

10	orf75-1.pep	MFQKHLQKASDSVVGGTLYVVATPIGNLADITRLALAVLQKADIICAEDTRVTAQLLSAY
	orf75ng-1	MFQKHLQKASDSVVGGTLYVVATPIGNLADITRLALAVLQKADIICAEDTRVTAQLLSAY
15	orf75-1.pep	GIQGLVSVREHNERQMA DKIVGYLS DGMVVAQVSDAGT PAVCDPGAKLARRVREAGFKV
	orf75ng-1	GIQGLVSVREHNERQMA DKIVGYLS DGLVVAQVSDAGT PAVCDPGAKLARRVREAGFKV
20	orf75-1.pep	VPVVGASAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLF AKWVRAAFVVMFETPHRIG
	orf75ng-1	VPVVGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLF AKWVRAAFVVMFETPHRIG
25	orf75-1.pep	ATLADMAELFPERRMLLAREITKTFFETFLSGTVGEIQTALSADGNQSRGEMVLVLYPAQD
	orf75ng-1	ATLADMAELFPERRMLLAREITKTFFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQD
30	orf75-1.pep	EKHEGLSESAQNIMKILTAELPTKQAAELA AKITGEGKKALYDLALSWKNKX
	orf75ng-1	EKHEGLSESAQNAMKILAAELPTKQAAELA AKITGEGKKALYDLALSWKNKX

Furthermore, ORG75ng-1 shows significant homology to a hypothetical *E.coli* protein:

sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION (F286)
 40 >gi|606086 (U18997) ORF f286 [Escherichia coli]
 >gi|1789535 (AE000395) Hypothetical 31.3 kD protein in agai-mtr intergenic region [Escherichia coli] Length = 286
 Score = 218 bits (550), Expect = 3e-56
 Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)
 45 Query: 4 KHLQKASDSVVGGTLYVVATPIGNLADITRLALAVLQKADIICAEDTRVTAQLLSAYGIQ 63
 K Q A +S G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI
 Sbjct: 2 KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLQAVDLIAAEDTRHTGLLQHFGIN 59
 50 Query: 64 GRLVSVREHNERQMA DKVIGFLSDGLVVAQVSDAGT PAVCDPGAKLARRVREAGFKVVPV 123
 RL ++ +HNE+Q A+ ++ L +G +A VSDAGT P + DPG L R REAG +VVP+
 Sbjct: 60 ARLFALHDHNEQQKAETLLAKLQEQNIALVSDAGTFLINDPGYHLVRTCREAGIRVVPL 119
 55 Query: 124 VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLF AKWVRAAFVVMFETPHRIGATL 183
 G A + ALS AG+ F + GF+P KS RR ++ +E+ HR+ +L
 Sbjct: 120 PGPCAAITLSAAGLPSDRFCYEGFLPAKSKGRRDALKAEAEPTLIFYESTHRLDLSL 179
 Query: 184 ADMAELFPERR-LMLAREITKTFFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242
 D+ + E R ++LARE+TKT+ET VGE+ + D N+ +GEMVL++ +
 60 Sbjct: 180 EDIVAVLGESRYVVLARELTKTWETIHGAPVGELLAWVKEDENRRKGEMVLIV-EGHKAQ 238
 Query: 243 HEGLSESAQNAMKILAAELPTKQAAELA AKITGEGKKALYDLAL 286
 E L A + +L AELP K+AA LAA+I G K ALY AL
 65 Sbjct: 239 EEDLPADALRTLALLQAEPLKKAALAAEIHGVKKNALYKYAL 282

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 35

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 295>:

```

1 ATGAAACAGA AAAAACCAGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
51 TTTTGGCGCA GC.AAAGCAC CCGAAATCGA CCCGGCTTTG .....
//
651 ..... .GAGTTGG TCAGAAACCA GTTGGAGCAG GGTTTGAGAC
10 701 AGGAAAAAGC CCGCTTGAAA ATCGATGCC TTTTGAAGA AAACGGTGTG
751 AAACGTAA

```

This corresponds to the amino acid sequence <SEQ ID 296; ORF76>:

```

1 MKQKKTAAAV IAAMLAGFAA XKAPEIDPAL .....
//
15 201 ..... ELVRNQLEQG LRQEKARLKI DALLEENGVK
251 P*

```

Further work revealed the complete nucleotide sequence <SEQ ID 297>:

```

1 ATGAAACAGA AAAAACCAGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
51 TTTTGGCGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
20 101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGCC GGCTACAAAC
201 TTTGGAAGTT TTGAAAAACA GGCATTGAA GGAAGGTTG GATAAGGATA
251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
301 GAGTACGTCC GTTTTCTGGA ACGTTCCGAA ACGGTTTCCG AAGACGAGCT
25 351 GCACAAGTTT TACGAACAGC AAATCCGCAT GATCAAATG CAGCAGGTCA
401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACAGCAGGC
501 TTTTGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTCGC
551 AGTTTGCCGC GATGAATCGG GGCAGCTTA CCCGCGATCC GGTCAAATTG
30 601 GGCGAACGCT ATTATCTGTT CAAACTCAGC GAGGTCGGGA AAAACCCCGA
651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGGAGCAG GGTTTGAGAC
701 AGGAAAAAGC CCGCTTGAAA ATCGATGCC TTTTGAAGA AAACGGTGTG
751 AAACGTAA

```

This corresponds to the amino acid sequence <SEQ ID 298; ORF76-1>:

```

35 1 MKQKKTAAAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
51 KPDGQAIRND AVRRLQTLV LKNRALKEGL DKDKDVQNR KIAEASFYAE
101 EYVRFLESE TVSEDELHKF YEQQIRMIKL QQVSFATEEE ARQAQQLLLK
151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAAMNR GDVTRDPVKL
40 201 GERYYLFLKLS EVGKNPDAQP FELVRNQLEQ GLRQEKARLK IDALLEENG
251 KP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF76 shows 96.7% identity over a 30aa overlap and 96.8% identity over a 31aa overlap with an ORF (ORF76a) from strain A of *N. meningitidis*:

```

45 orf76.pep      10      20      30
      MKQKKTAAAVIAAMLAGFAAKKAPEIDPAL
      |||||
orf76a      10      20      30      40      50      60
      MKQKKTAAAVIAAMLAGFAAKKAPEIDPALVDTLVAQIMQQADRHAEQSQKPDGQAIRND
      10      20      30      40      50      60
//
70      80      90

```


orf76.pep XELVRNQLQGLRQEKARLKIDALLEENGVPKX
 orf76a DVTRDPVKLGERYYLFLKSEVGKNPDAQPFELVRNQLQGLRQEKARLKIDALLEENGVPKX
 200 210 220 230 240 250

5 The complete length ORF76a nucleotide sequence <SEQ ID 299> is:

1 ATGAAACAGA AAAAAACCGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
 51 TTTTGGCGCA GCCAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
 101 TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
 151 AAACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGTC GGCTGCAAAAC
 201 TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
 251 AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
 301 GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCCG AAAGCGCACT
 351 GCGTCAGTTT TATGAGCGGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
 401 GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
 451 GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
 501 TTTTGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTCGC
 551 AGTTTGACGC GATGAATCGG GCGCAGCTTA CCCGCGATCC GGTCAAATTG
 601 GCGCAACGCT ATTATCTGTT CAACTCAGC GAGGTCGGGA AAAACCCCGA
 651 CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGGAAACA GGTGTGAGAC
 701 AGGAAAAAGC CCGCTTGAAA ATCGATGCCA TTTTGAAGA AAACGGGTGC
 751 AAACCGTAA

This encodes a protein having amino acid sequence <SEQ ID 300>:

1 MKQKKTAAAV IAAMLAGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAEQSQ
 51 KPDGQAIRND AVRRLQTLV LKNRALKEGL DKDKDVQNRK KIAEASFYAE
 101 EYVRFLESE TVSESALRQF YERQIRMIKL QQVSFATEEE ARQAQQLLLK
 151 GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAAMNR GDVTRDPVKL
 201 GERYYLFLKS EVGKNPDAQP FELVRNQLQ GLRQEKARLK IDAILEENGV
 251 KP*

ORF76a and ORF76-1 show 97.6% identity in 252 aa overlap:

30 orf76a.pep 10 20 30 40 50 60
 orf76-1 10 20 30 40 50 60
 35 orf76a.pep 70 80 90 100 110 120
 orf76-1 70 80 90 100 110 120
 40 orf76a.pep 130 140 150 160 170 180
 orf76-1 130 140 150 160 170 180
 45 orf76a.pep 190 200 210 220 230 240
 orf76-1 190 200 210 220 230 240
 50 orf76a.pep 250
 orf76-1 250

60 Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF76 and a predicted ORF (ORF76.ng) from *N. gonorrhoeae* of the N- and C-termini show 96.7 % and 100% identity in 30 and 31 overlap, respectively:

```

orf76.pep      MKQKKTAAAVIAAMLGFAAXKAPEIDPAL      30
                |||||
orf76ng        MKQKKTAAAVIAAMLGFAAAKAPEIDPALVDTLVAQIMQADRHAESQSRPDGQAIRND  60
                //
5  orf76.pep      ELVRNQLEQGLRQEKARLKIDALLEENGVKP  251
                |||||
orf76ng        VTRNPVKLGERYYLFKLGA VGKNPDAQPFELVRNQLEQGLRQEKARLKIDALLEENGVKP  251

```

The complete length ORF76ng nucleotide sequence <SEQ ID 301> is:

```

10      1  ATGAAACAGA AAAAGACCGC TGCCGCAGTT ATTGCTGCAA TGTTGGCAGG
      51  TTTTGCGGCA GCCAAAGCAC CCGAAATCGA CCCGGCTTTG GTGGATACGC
     101  TGGTGGCGCA GATCATGCAG CAGGCAGACC GGCATGCGGA GCAGTCCCAA
     151  AGACCGGACG GGCAGGCAAT CCGAAACGAT GCCGTCCGCC GGCTGCAAAAC
     201  TTTGGAAGTT TTGAAAAACA GGGCATTGAA GGAAGGTTTG GATAAGGATA
     251  AGGATGTCCA AAACCGCTTT AAAATCGCCG AAGCGTCTTT TTATGCCGAG
     301  GAGTACGTCC GTTTTCTGGA ACGTTCGGAA ACGGTTTCGG AAAGCGCACT
     351  GCGTCAGTTT TATGAGCGGC AAATCCGCAT GATCAAATTG CAGCAGGTCA
     401  GCTTCGCAAC CGAAGAGGAG GCGCGTCAGG CGCAGCAGCT CCTGCTCAAA
     451  GGGCTGTCTT TTGAAGGGCT GATGAAGCGT TATCCGAACG ACGAGCAGGC
     501  GTTCGACGGT TTCATTATGG CGCAGCAGCT TCCCGAGCCG CTGGCTTcgc
     551  agtttgcgg TATGAACCGT GCGCAGCTTA CCCGCAATCC GGTCAAATTG
     601  GGCGAACGCT ATTACCTGTT CAAACTCGGC GCGGTCGGGA AAAACCCCGA
     651  CGCGCAGCCT TTCGAGTTGG TCAGAAACCA GTTGAACAA GGTGTGAGGC
     701  AGGAAAAAGC CCGCTTGAAA ATCGATGCCC TTTTGAaga Aaacggtgtc
     751  AaacCGTAA

```

25 This encodes a protein having amino acid sequence <SEQ ID 302>:

```

1  MKQKKTAAAV IAAMLGFAA AKAPEIDPAL VDTLVAQIMQ QADRHAESQ
51  RPDGQAIRND AVRRLQTLV LKNRALKEGL DKDKDVQNRF KIAEASFYAE
101  EYVRFLESE TVSESALRQF YERQIRMIKL QQVSFATEEE ARQAQQLLLK
151  GLSFEGLMKR YPNDEQAFDG FIMAQQLPEP LASQFAGMNR GDVTRNPVKL
201  GERYYLFLKG AVGKNPDAQP FELVRNQLEQ GLRQEKARLK IDALLEENG
251  KP*

```

ORF76ng and ORF76-1 show 96.0% identity in 252 aa overlap

```

35      10      20      30      40      50      60
orf76-1.pep  MKQKKTAAAVIAAMLGFAAAKAPEIDPALVDTLVAQIMQADRHAESQKPDGQAIRND
                |||||
orf76ng      MKQKKTAAAVIAAMLGFAAAKAPEIDPALVDTLVAQIMQADRHAESQSRPDGQAIRND
                |||||
      10      20      30      40      50      60
40  orf76-1.pep  AVRRLQTLV LKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLESETVSEDELHKF
                |||||
orf76ng      AVRRLQTLV LKNRALKEGLDKDKDVQNRFKIAEASFYAEYVRFLESETVSESALRQF
                |||||
      70      80      90      100     110     120
45  orf76-1.pep  YEQQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
                |||||
orf76ng      YERQIRMIKLQQVSFATEEEARQAQQLLLKGLSFEGLMKRYPNDEQAFDGFIMAQQLPEP
                |||||
      130     140     150     160     170     180
50  orf76-1.pep  LASQFAAMNRGDVTRDPVKLGERYYLFKLSEVGKNPDAQPFELVRNQLEQGLRQEKARLK
                |||||
orf76ng      LASQFAGMNRGDVTRNPVKLGERYYLFKLGA VGKNPDAQPFELVRNQLEQGLRQEKARLK
                |||||
      190     200     210     220     230     240
55  orf76-1.pep  IDALLEENGVPKX
                |||||
orf76ng      IDALLEENGVPKX
                |||||
      250
60

```

Furthermore, ORF76ng shows significant homology to a *B.subtilis* export protein precursor:

sp|P24327|PRSA_BACSU PROTEIN EXPORT PROTEIN PRSA PRECURSOR >gi|98227|pir||S15269
 33K lipoprotein - Bacillus subtilis >gi|39782 (X57271) 33kDa lipoprotein
 [Bacillus subtilis]
 >gi|2226124|gnl|PID|e325181 (Y14077) 33kDa lipoprotein [Bacillus subtilis]
 5 >gi|2633331|gnl|PID|e1182997 (Z99109) molecular chaperonin [Bacillus subtilis]
 Length = 292
 Score = 50.4 bits (118), Expect = 1e-05
 Identities = 48/199 (24%), Positives = 82/199 (41%), Gaps = 32/199 (16%)

10 Query: 70 VLKNRALKEGLDK-----DKDVQNRFKIAEASF-----YAEYVRFLESETVSE 114
 VL ++ LDK DK++ N+ K + Y ++Y++ + E +++
 Sbjct: 53 VLTQLVQEKVLDKKYKVSDEIDNKLKEYKTQLGDQYTALEKQYKGKDYKEQVKYELLTQ 112

15 Query: 115 SA-----LRQFYERQIRMILQOVSFATEEEARQAQQLLLKGLSFEGLMKRYPN 163
 A +++++E I+ + A ++ A + ++ L KG FE L K Y
 Sbjct: 113 KAAKDNIKVTDAIDKEYWEGKLGKIRASHILVADKKTAEVEKKLKGKGFEDLAKEYST 172

20 Query: 164 DEQAFDG-----FIMAQQLPEPLASQFAAMNRGDVTRDPVKLGERYYLKFLSEVGKPNDA 218
 D A G F Q+ E + + G+V+ DPVK Y++ K +E D
 Sbjct: 173 DSSASKGGDLGWFAKEGQMDTFFSKAAFKLKTGEVS-DPVKTQYGYHIIKKTTEERGKYDD 231

25 Query: 219 QPFELVRNLEQGLRQEKA 237
 EL LEQ L A
 Sbjct: 232 MKKELKSEVLEQKLNDNAA 250

Based on this analysis, including the presence of a putative leader sequence and a RGD motif in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF76-1 (27.8kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 10A shows the results of affinity purification of the His-fusion protein, Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 10B), ELISA (positive result), and FACS analysis (Figure 10C). These experiments confirm that ORF76-1 is a surface-exposed protein, and that it is a useful immunogen.

35 Example 36

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 303>:

1 ATGAAAAAAT CTTTCCTTAC GCTTGTCTG TATTCGTCTT TACTTACCGC
 51 CAGCGAAATT GCCTTACCCC TTGGAATTGG GGATTGAAAC CTTACCGGCG
 101 GCAAAAATTG CGGAAACGTT TGCCTGACA TTTGTGATTG CTGCGCTGTA
 140 151 TCTGTTTGGC CGTAATAAGG TGACGCGTTT GTTGATTGCG GTGTTTTTTG
 201 CGTTCAGCAT TATTGCCAAC AATGTGCATT ACGCGGATTA TCAAAGCTGG
 251 ATGACG....
 //

1201 CAAACCGTAT TCGAGCAGCT GCAAAAGACT CCTGACGGCA
 145 1251 ACTGGCTGTT TGCCTATACC TCCGATCATG GCCAGTATGT TCGCCAAGAT
 1301 ATCTACAATC AAGGCACGGT GCAGCCCGAC AGCTATCTCG TGCCGCTAGT
 1351 GTTGTACAGC CCGGATAAGG CCGTGCAACA GGCTGCCAAC CAGGCTTTTG
 1401 CGCCTTGCGA GATTGCCTTC CATCAGCAGC TTTCAACGTT CCTGATTACAC
 1451 ACGTTGGGCT ACGATATGCC GGTTCAGGT TGTCCGGAAG GCTCGGTAAC
 50 1501 GGGCAACCTG ATTACGGGTG ATGCAGGCAG CTTGAACATT CGCGACGGCA
 1551 AGGCGGAATA TGTTTATCCG CAATGA

This corresponds to the amino acid sequence <SEQ ID 304; ORF81>:

1 MKKSFLTLVL YSSLLTASEI AYPLELGIET LPAKIAETF ALTFVIAALY

Further work revealed the complete nucleotide sequence <SEQ ID 305>:

40

45

Homology with a predicted ORF from *N.meningitidis* (strain A)

55

60

```

orf81.pep      LIAVFFAFSIIANNVHYADYQSWMT
                |||||
orf81a         LIAVFFAFSIIANNVHYAVYQSWITGINYWLMLEITEVGGAGASMLDKLWLPALWGVLE
                70      80      90      100     110     120
                //
orf81.pep      QTVFEQLQKTPDGNWLFAYTSDHGQYVRQD
                |||||
orf81a         IPHANGLEQISGGDIVDKYDNTIHKTDOMIQTVFEQLQKQPDGNWLFAYTSDHGQYVRQD
                280     290     300     310     320     330
orf81.pep      IYNQGTVPQDSYLVPLVLVYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYPVSG
                |||||
orf81a         IYNQGTVPQDSYLVPLVLVYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYPVSG
                340     350     360     370     380     390
orf81.pep      CREGSVTGNLITGDAGSLNIRDGKAEYVYPQX
                |||||
orf81a         CREGSVTGNLITGDAGSLNIRDGKAEYVYPQX
                400     410     420

```

	25	1	ATGAAAAAAT	CCCTTTTCTG	TCTCTTTCTG	TATTCTGCCC	TACTTACTGC
		51	CAGCGAAATT	GCTTATCGTT	TTGTATTCTG	AATTGAAACC	TTACCGGCTG
		101	CAAAAATGGC	AGAAACGTTT	GCGCTGACAT	TTGTGATTGC	TGCGCTGTAT
		151	CTGTTTGC	GTTATAAGGC	AACGCGTTTG	TTGATTGCGG	TGCTTTTTCGC
		201	GTTCAGCATT	ATTGCCAACA	ATGTGCATTA	CGCGGTTTAT	CAAAGCTGGG
		251	TAACGGGCAT	TAATTATTGG	CTGATGCTGA	AAGAGATTAC	CGAAGTTGGC
	30	301	GGCGCAGGGG	CGTCGATGTT	GGATAAGTTG	TGGTGCTCTG	CGCTGTGGGG
		351	CGTGTTGGAA	GTCGTGTTGT	TTTGACAGCT	TGCCAAGTTC	CGCCGTAAGA
		401	CGCATTTTTC	TGCCGATATA	CTGTTTGCTC	TCCTAATGCT	GATGATTTTC
		451	GTGCGTTCGT	TCGACACGAA	ACAAGAACAC	GGTATTTCGC	CCAAACCCGAC
		501	ATACAGCCGC	ATCAAAGCCA	ATTATTTCAG	CTTCGGTTAT	TTTGTCTGGAC
	35	551	CGGTGTTGCC	GTATCAGTTG	TTTGATTATA	GCAAGATTCC	TGTGTTCAAA
		601	CAGCCTGCTC	CAAGCAGAAT	CGGGCAAGGC	AGTATTCAAA	ATATCGTCTC
		651	GATTCTGGGC	GAAAGCGAAA	GCGCGGCGCA	TTTGAATTAG	TTTGCTACG
		701	GGCGCGAAAC	TTGCGCGTTT	TTGACCCAGC	TTTCGCAAGC	CGATTTTAAG
		751	CCGATTGTGA	AACAAAGTTA	TTCCGCAGGC	TTTATGACGG	CAGTATCCCT
	40	801	GCCCAGTTTC	TTTAACGTCA	TACCGCATGC	CAACGCGTTG	GAACAAATCA
		851	CGGCGGGCGA	TATTGTGGAT	AAGTACGACA	ACACCATCCA	CAAAACCGAC
		901	CAAAATGATT	AAACCGTATT	CGAGCAGCTG	CAAAAGCAGC	CTGACGGCAA
		951	CTGGCTGTTT	GCTATACCT	CCGATCATGG	CCAGTATGTT	CGCCAAGATA
		1001	TCTACAATCA	AGGCACGGTG	CAGCCCAGCA	GCTATCTCGT	GCCGCTGGTG
	45	1051	TTGTACAGCC	CGGATAAGGC	CGTGCAACAG	GCTGCAACAG	AGGCTTTTGC
		1101	GCCTTGCGAG	ATTGCCTTCC	ATCAGCAGCT	TTCAACGTTT	CTGATTACAA
		1151	CGTTGGGCTA	CGATATGCCG	GTTTCAGGTT	GTCGCGAAGG	CTCGGTAACG
		1201	GGCAACCTGA	TTACGGGTGA	TGCAGGCAGC	TTGAACATT	GCGACGGCAA
		1251	GCGCGAATAT	GTTTATCCGC	AATGA		

1 MKKSLFVLFL YSLLLTASEI AYRFVFGIET LPAAKMAETF ALTFFVIAALY
 51 LFARYKATRL LIAVFFAFSI IANNVHYAVY QSWITGINYW LMLKEITEVG
 101 GAGASMLDKL WLPALWGVLE VMLFCSLAKF RRKTHFSADI LFAFLMLMIF
 151 VRSFDTKQEH GISPKPTYSR IKANYFSFGY FVGRVLDPYQL FDLSKIPIPVFK
 201 QPAPSRIGQG SIQNIVLIMG ESESAHLKL FGYGRETS PF LTQLSQADFK
 251 PIVKQSYSAG FMTAVSLPSF FNVIPHANGI EQISGGDIVD KYDNTIHKTD
 301 QMIQTVFEQL QKQPDGNWLF AYTSDHGQYV RQDIYNQGTV QPDSYLVPLV
 351 LYSPPDKAVQ ANQAFAPCE IAFHQQLSTF LIHTLYDMP VSGCREGSVT
 401 GNLTGDAGS LNIRDGAKEY VYPQ*

```

                                10      20      30      40      50      60
orf81a.pep    MKKSLFVLFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALT FVIAALYLFARYKATRL
               ||||::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf81-1       MKKSFLT LVLYSSLLTASEIAYRFVFGIETLPAAKIAETFALT FVIAALYLFARYKVTRL
               10      20      30      40      50      60

```

-207-

		70	80	90	100	110	120
	orf81a.pep	LIAVFFAFSIIANNVHYAVYQSWITGINYWMLKEITEVGGAGASMLDKLWLPALWGVLE					
5	orf81-1	LIAVFFAFSIIANNVHYAVYQSWMTGINYWMLKEVTEVGSAGASMLDKLWLPVWGVLE					
		70	80	90	100	110	120
	orf81a.pep	VMLFCSLAKFRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY					
10	orf81-1	VMLFCSLAKFRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSRIKANYFSFGY					
		130	140	150	160	170	180
	orf81a.pep	FVGRVLPYQLFDLSKIPVFKQAPPSRIGQGSIQNIVLIMGESESAHLKLFGYGRETSPF					
15	orf81-1	FVGRVLPYQLFDLSRIPAFKQAPPSKIGQGSVQNIIVLIMGESESAHLKLFGYGRETSPF					
		190	200	210	220	230	240
	orf81a.pep	LTQLSQADFKPIVKQSYSAGFMTAVSLPSFFNVIPHANGLEQISGGD-----					
20	orf81-1	LTRLNQADFKPIVKQSYSAGFMTAVSLPSFFNAIPHANGLEQISGGDTNMFRLAKEQGYE					
		250	260	270	280	290	300
25	orf81a.pep	-----					
	orf81-1	TYFYSAQAENEMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQQGKHF					
30		310	320	330	340	350	360
	orf81a.pep	-----		290	300	310	320
				IVDKYDNTIHKTQMIQTVEQLQKQPDGNWLF			
35	orf81-1	IVLHQRGSHAPYGALLQPDQKVFGEADIVDKYDNTIHKTQMIQTVEQLQKQPDGNWLF					
		370	380	390	400	410	420
	orf81a.pep	AYTSDHGQYVRQDIYNQGTVPDPSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
40	orf81-1	AYTSDHGQYVRQDIYNQGTVPDPSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
		430	440	450	460	470	480
	orf81a.pep	LIHTLGYDMPVSGCREGVSVTGNLITGDAGSLNIRDGKAEYVYPQX					
45	orf81-1	LIHTLGYDMPVSGCREGVSVTGNLITGDAGSLNIRDGKAEYVYPQX					
		490	500	510	520		

50 Homology with a predicted ORF from *N.gonorrhoeae*

The aligned aa sequences of ORF81 and a predicted ORF (ORF81.ng) from *N. gonorrhoeae* of the N- and C-termini show 82.4 % and 97.5% identity in 85 and 121 overlap, respectively:

	orf81.pep	MKKSFLTLVLYSSLLTASEIAYPLELGIETLPAAKIAETFALT FVIAALYLFARNKVTRL	60
55	orf81ng	MKKSFLVLFYSSLLTASEIAYRFVGIETLPAAKMAETFALT FVIAALYLFARYKASRL	60
	orf81.pep	LIAVFFAFSIIANNVHYADYQSWMT	85
	orf81ng	LIAVFFAFSMIANNVHYAVYQSWMTGINYWMLKEVTEVGSAGASMLDKLWLPALWGVAE	120
60	orf81.pep	QTVEQLQKTPDGNWLFAYTSDHGQYVRQD	433
	orf81ng	ALLQPDQKVFGEADIVDKYDNTIHKTQMIQTVEQLQKQPDGNWLFAYTSDHGQYVRQD	433
65	orf81.pep	IYNQGTVPDPSYLVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG	493
	orf81ng	IYNQGTVPDPSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTFLIHTLGYDMPVSG	493

-208-

```

orf81.pep    CREGSVTGNLITGDAGSLNIRDGKAEYVYPQ  524
             |||||
orf81ng      CREGSVTGNLITGDAGSLNIRNGKAEYVYPQ  524

```

The complete length ORF81ng nucleotide sequence <SEQ ID 309> is:

```

5      1  ATGAAAAAAT CCCTTTTCGT TCTCTTCTG TATTCATCCC TACTTACCGC
      51  CAGCGAAATC GCCTATCGCT TTGTATTCGG AATTGAAACC TTACCGGCTG
     101  CAAAAATGGC GGAAACGTTT GCGCTGACAT TTATGATTGC TGCGCTGTAT
     151  CTGTTTGCGC GTTATAAGGC TTCGCGGCTG CTGATTGCGG TGTTTTTCGC
     201  GTTCAGCATG ATTGCCAACA ATGTGCATTA CGCGGTTTAT CAAAGCTGGA
     251  TGACGGGTAT TAACTATTGG CTGATGCTGA AAGAGGTTAC CGAAGTCGGC
     301  AGCGCGGGCG CGTCGATGTT GGATAAGTTG TGGCTGCCTG CTTTGTGGGG
     351  CGTGGCGGAA GTCATGTTGT TTTGCAGCCT TGCCAAGTTC CGCCGTAAGA
     401  CGCATTTTTC TGCCGATATA CTGTTTGCCT TCCTAATGCT GATGATTTC
     451  GTGCGTTCGT TCGACACGAA ACAAGAGCAC GGTATTTCCG CCAAACCGAC
     501  ATACAGCCGC ATCAAAGCCA ATTATTTTCT CTTCGGTTAT TTTGTCGGGC
     551  GCGTGTGCC GTATCAGTTG TTTGATTAA GCAAGATCCC TGTGTTCAAA
     601  CAGCCTGCTC CAAGCAAAAT CGGGCAAGGC AGTATTCAAA ATATCGTCTT
     651  GATTATGGGC GAAAGCGAAA GCGCGGCGCA TTTGAAATTG TTTGGTTACG
     701  GCGCGGAAAC TTCGCCGTTT TTAACCCGGC TGTCGCAAGC CGATTTTAAG
     751  CCGATTGTGA AACAAAGTTA TTCCGCAGGC TTTATGACGG CAGTATCCCT
     801  GCCCAGTTTC TTTAACGTCA TACCGCACGC CAACGGCTTG GAACAAATCA
     851  GCGCGGCGCA TACCAATATG TTCCGCCTCG CCAAAGAGCA GGGCTATGAA
     901  ACGTATTTTT ACAGTGCCCA GGCTGAAAAC CAAATGGCAA TTTTGAAGTT
     951  AATCGGTAAG AAATGGATAG ACCATCTGAT TCAGCCGACG CAACTTGGCT
    1001  ACGGCAACGG CGACAATATG CCCGATGAGA AGCTGCTGCC GTTGTTCGAC
    1051  AAAATCAATT TGCAGCAGGG CAGGCATTTT ATCGTGTTCG ACCAACGCGG
    1101  TTCGCACGCC CCATACGGCG CATTTGTGCA GCCTCAAGAT AAAGTATTCG
    1151  GCGAAGCCGA TATTGTGGAT AAGTACGACA ACACCATCCA CAAAACCGAC
    1201  CAAATGATTG AAACCGTATT CGAGCAGCTG CAAAAGCAGC CTGACGGCAA
    1251  CTGGCTGTTT GCCTATACCT CCGATCATGG CCAGTATGTG CGCCAAGATA
    1301  TCTACAATCA AGGCACGGTG CAGCCCGACA GCTATATTGT GCCTCTGGTT
    1351  TTGTACAGCC CGGATAAGGC CGTGCAACAG GCTGCCAACC AGGCTTTTGC
    1401  GCCTTGCGAG ATTGCCTTCC ATCAGCAGCT TTCAACGTTT CTGATTACAA
    1451  CGTTGGGCTA CGATATGCCG GTTTCAGGTT GTCGCGAAGG CTCGGTAACA
    1501  GGCAACCTGA TTACGGGCGA TGCAGGCAGC TTGAACATTC GCAACGGCAA
    1551  GGCGGAATAT GTTTATCCGC AATAA

```

This encodes a protein having amino acid sequence <SEQ ID 310>:

```

      1  MKKSLFVLFL YSLLTASEI AYRFVFGIET LPAAKMAETF ALTFMIAALY
     51  LFARYKASRL LIAVFFAFSM IANNVHYAVY QSWMTGINYW LMLKEVTEVG
    101  SAGASMLDKL WLPALWGVAE VMLFCSLAKF RRKTHFSADI LFAFLMLMIF
    151  VRSFDTKQEH GISPKPTYSR IKANYFSFGY FVGRVLPYQL FDLISKIPVK
    201  QPAPSKIGQG SIQNIVLIMG ESESAHLKL FGYGRETSPF LTRLAQADFK
    251  PIVKQSYSAG FMTAVSLPSF FNVIPHANGL EQISGGDTNM FRLAKEQGYE
    301  TYFYSAQAEQ QMAILNLIGK KWIDHLIQPT QLGYNNGDNM PDEKLLPLFD
    351  KINLQQRHF IVLHQGRSHA PYGALLQPD KVFGEADIVD KYDNTIHKTD
    401  QMIQTVFEQL QKQPDGNWLF AYTSDHGQYV RQDIYNQGTQ QPDSYIVPLV
    451  LYSPDKAVQQ AANQAFAPCE IAFHQQLSTF LIHTLGYDMP VSGCREGSVT
    501  GNLITGDAGS LNIRNGKAEY VYPQ*

```

ORF81ng and ORF81-1 show 96.4% identity in 524 aa overlap:

```

50      10      20      30      40      50      60
orf81ng-1.pep MKKSLFVLFLYSSLLTASEIAYRFVFGIETLPAAKMAETFALTFMIAALYLFARYKASRL
             |||||
orf81-1        MKKSFLTLYSSLLTASEIAYRFVFGIETLPAAKIAETFALT FVIAALYLFARYKVTRL
             10      20      30      40      50      60

55      70      80      90      100     110     120
orf81ng-1.pep LIAVFFAFSMIANNVHYAVYQSWMTGINYWMLKEVTEVGSAGASMLDKLWLPALWGVAE
             |||||
orf81-1        LIAVFFAFSIIANNVHYAVYQSWMTGINYWMLKEVTEVGSAGASMLDKLWLPVWLVLE
             70      80      90      100     110     120

60      130     140     150     160     170     180
orf81ng-1.pep VMLFCSLAKFRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSR IKANYFSFGY
             |||||
orf81-1        VMLFCSLAKFRKTHFSADILFAFLMLMIFVRSFDTKQEHGISPKPTYSR IKANYFSFGY

```

-209-

		130	140	150	160	170	180
		190	200	210	220	230	240
5	orf81ng-1.pep	FVGRVLPYQLFDLSKIPVFKQPAPSKIGQGSIQNIVLIMGESESAHLKLFYGRGRETSPF					
	orf81-1	FVGRVLPYQLFDLSRIPAFKQPAPSKIGQGSVQNIIVLIMGESESAHLKLFYGRGRETSPF					
		190	200	210	220	230	240
10	orf81ng-1.pep	LTRLQADFKPIVKQSYSAGFMTAVSLPSFFNVI PHANGLEQISGGDTNMFR LAKEQGYE					
	orf81-1	LTRLQADFKPIVKQSYSAGFMTAVSLPSFFNAI PHANGLEQISGGDTNMFR LAKEQGYE					
		250	260	270	280	290	300
15	orf81ng-1.pep	TYFYSAQAENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQOGRHF					
	orf81-1	TYFYSAQAENEMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQOGRHF					
		310	320	330	340	350	360
20	orf81ng-1.pep	IVLHQRGSHAPYGALLQPDQKVFGEADIVDKYDNTIHKTDQMIQT VFEQLQKQPDGNWLF					
	orf81-1	IVLHQRGSHAPYGALLQPDQKVFGEADIVDKYDNTIHKTDQMIQT VFEQLQKQPDGNWLF					
25		370	380	390	400	410	420
	orf81ng-1.pep	AYTSDHGQYVRQDIYNQGTVPDQSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
	orf81-1	AYTSDHGQYVRQDIYNQGTVPDQSYIVPLVLYSPDKAVQQAANQAFAPCEIAFHQQLSTF					
		430	440	450	460	470	480
30	orf81ng-1.pep	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRNGKAEYVYPQX					
	orf81-1	LIHTLGYDMPVSGCREGSVTGNLITGDAGSLNIRDGKAEYVYPQX					
		490	500	510	520		
35	orf81ng-1.pep						
	orf81-1						

Furthermore, ORF81ng shows significant homology to an *E. coli* OMP:

40	gil1256380 (U50906) outer membrane adherence protein-associated protein [E. coli] Length = 547 Score = 87.4 bits (213), Expect = 2e-16 Identities = 122/468 (26%), Positives = 198/468 (42%), Gaps = 70/468 (14%)
45	Query: 25 VFGIETLPAAKMAETFA-LTFMIAALYLFARYKAS--RLLIAVFFAFSMIANNVHYAVYQ 81 VFGI L A+ A L F + + + R + RLL+A F + A ++ ++Y Sbjct: 29 VFGITNLVASSGAHMVQRLLFFVLTLIVVKRISSLPLRLLVAAPEVL-LTAADMSISLY- 86
50	Query: 82 SWMT-----GINYWMLKEVTEVGSAGASMLDKLWLPALWGVAEVMFLCSLAKFRKRT 134 SW T G ++ + EV A ML ++ P L A + L + Sbjct: 87 SWCTFGTTFNDGFAISVLQSDPDEV----AKMLG-MYSPYLCAFAFLSLLFLAVIICYDV 141
55	Query: 135 HFSADILFAFLMLMIFVRSF-----DTKQEHGISPKPTYSRIKAN--YFSFGYFVG 183 + L+L++ S D K ++ SP SR +F+ YF Sbjct: 142 SLPTKKVTGILLLIVISGSLFSACQFAYKDAKNKNAFSPYILASRFATYTPFFNLNYFAL 201
60	Query: 184 RVLPHYQ--LFDLSKIPVFKQPAPSKIGQGSIQNIVLIMGESESAHLKLFYGRGRETSPFL 241 +Q L + +P F+ + I VLI+GES ++ L+GY R T+P + Sbjct: 202 AAKEHQRLLSIANTVPYFQL----SVRDTGIDTYVLIVGESVRVDNMSLYGYTRSTTPQV 257
65	Query: 242 TRLSQADFKPIVKQSYSAGFMTAVSLP---SFFNVI PHANGLEQISGGDTNMFR LAKEQG 298 +Q + Q+ S TA+S+P + +V+ H I N+ +A + G Sbjct: 258 E--AQRKQIKLFNQAISGAPYTALSVPPLSLTADSVLSH-----DIHNYPDNINMANQAG 310
70	Query: 299 YETIFYSAQA---ENQMAILNLIGKKWIDHLIQPTQLGYGNGDNMPDEKLLPLFDKINLQ 355 ++T++ S+Q+ +N A+ ++ ++ + Y G DE LLP + Q Sbjct: 311 FQTFWLSSQSAFRQNGTAVTSI-----AMRAMETVYVRGF---DELLPLHLSQALQQ 359
	Query: 356 --QGRHFIVLHQRGSHAPYGALLQPDQKVFGEADIVDK-YDNTIHKTDQMIQT VFEQLQK 412 Q + IVLH GSH P + VF D D YDN+IH TD ++ VFE L+ Sbjct: 360 NTQQKKLIVLHLNGSHEPACSAYPQSSAVFQPDQDDACYDNSIHYTDSL LGQVFELK- 418

Query: 413 QPDGNWLFAYTSDHG---QYVRQDIYNQG--TVQPD SYIVPL-VLYSP 454
 D Y +DHG ++++Y G +Y VP+ + YSP
 Sbjct: 419 --DRRASVMYFADHGLERDPTKKNVYFHGGREASQQAYHVMFIWYSP 464

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 37

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 311>:

```

1   ...ACCCGTGCTCC TCTTCATCCC CCTCGTCCTC ACAC.GTGGC GCACACTGAC
51  CGGCATACTC GCCCaCGGCG GCGGCAAACG CTTTGCCGTC GAACAAGAAC
101 TCGTCGCCGC ATCGTCCCGC GCCGCCGTCA AAGAAATGGA TTGTCCGCC
15  151 yTAAAGGAC GCAAGCCGC CyTTACGTC TCCGTTATGG GCGACCAAGG
201 TTCGGGCAAC ATAAGCGGCG GACGCTACTC TATCGACGCA CTGATACGCG
251 GCGGCTACCA CAACAACCCC GAAAGTGCCA CCCAATACAG CTACCCCGCC
301 TACGACACTA CCGCCACCAC CAAATCCGAC GCGCTCTCCA GCCTAACCC
351 TTCCACATCG CTTTGAACG CCCCCGCGC CGyCytGACG AAAAACAGCG
20  401 GACGCAAAGG CGAACGcTCC GCCGGACTGT CCGTCAACGG CACGGGCGAC
451 TACCGCAACG AAACCCTGCT CGCCAACCCC CGCGACGTTT CCTTCCTGAC
501 CAACCTCATC CAAACCGTCT TCTACCTGCG CGGCATCGAA GTCgTACCGC
551 CCGrATACGC CGACACCGAC GTATTCGTAA CCGTCGACGT A...
```

This corresponds to the amino acid sequence <SEQ ID 312; ORF83>:

```

25  1 ..TLLFIPLVL TXCGTLTGIL AHGGGKRFAV EQELVAASSR AAVKEMDLA
51  LKGRKAAXYV SVMGDQSGN ISGGRYSIDA LIRGGYHNNP ESATQSYPA
101 YDTTATTKSD ALSSVTTSTS LLNAPAXLT KNSGRKGRS AGLSVNGTGD
151 YRNETLLANP RDVSFLTNLI QTVFYLRGIE VVPPXYADTD VFVTVDV..
```

Further work revealed the complete nucleotide sequence <SEQ ID 313>:

```

30  1 ATGAAAACCC TGCTCCTCCT CATCCCCCTC GTCCTCACAG CCTGCGGCAC
51  ACTGACCGGC ATACCCGCCC ACGGCGGCGG CAAACGCTTT GCCGTGGAAC
101 AAGAACTCGT CGCCGCATCG TCCGCGCGCG CCGTCAAAGA AATGGATTGT
151 TCCGCCCTAA AAGGACGCAA AGCCGCCCTT TACGTCTCCG TTATGGGCGA
201 CCAAGGTTTCG GGCAACATAA GCGGCGGACG CTACTCTATC GACGCACTGA
35  251 TACGCGGCGG CTACCACAAC AACCCTGAAA GTGCCACCCA ATACAGCTAC
301 CCCGCCCTACG ACACTACCGC CACCACCAA TCCGACGCGC TCTCCAGCGT
351 AACCCTTCC ACATCGCTTT TGAACGCCCC CGCCGCCGCC CTGACGAAAA
401 ACAGCGGACG CAAAGCGGAA CGTCCGCCG GACTGTCCGT CAACGGCAG
451 GCGGACTACC GCAACGAAAC CTGTCTCGCC AACCCTCGCG ACGTTTCCTT
40  501 CCTGACCAAC CTCATCCAAA CCGTCTTCTA CCTGCGCGGC ATCGAAGTCG
551 TACCGCCCGA ATACGCCGAC ACCGACGTAT TCGTAACCGT CGACGTATTC
601 GGCACCGTCC GCAGCCGTAC CGAAGTGCAC CTCTACAACG CCGAAACCTT
651 TAAAGCCCAA ACCAAGCTCG AATATTTCGC CGTTGACCGC GACAGCCGGA
701 AACTGCTGAT TACCCCTAAA ACCGCCGCCT ACGAATCCCA ATACCAAGAA
45  751 CAATACGCCC TTTGGACCGG CCCTTACAAA GTCAGCAAAA CCGTCAAAGC
801 CTCAGACCGC CTGATGGTCG ATTTCTCCGA CATTACCCCC TACGGCGACA
851 CAACCGCCCA AAACCGTCCC GACTTCAAA AAAACAACGG TAAAAAACCC
901 GATGTCGGCA ACGAAGTCAT CCGCCGCCGC AAAGGAGGAT AA
```

This corresponds to the amino acid sequence <SEQ ID 314; ORF83-1>:

```

50  1 MKTLLLLIPL VLTACGTLTG IPAHGGGKRF AVEQELVAAS SRAAVKEMDL
51  SALKGRKAAL YVSMGDQGS GNISGGRYSI DALIRGGYHN NPESATQYSY
101 PAYDTTATTK SDALSSVTT TSLNAPAAA LTKNSGRKGE RSAGLSVNGT
151 GDYRNETLLA NPRDVSFLTN LIQTVFYLRG IEVVPPEYAD TDVFVTVDVF
201 GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLITPK TAAYESQYQE
55  251 QYALWTGPYK VSKTVKASDR LMFDFSDITP YGDTTAQNRP DFKQNNKKPK
```

301 DVGNEVIRRR KGG*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)ORF83 shows 96.4% identity over a 197aa overlap with an ORF (ORF83a) from strain A of *N.*5 *meningitidis*:

		10	20	30	40	50
orf83.pep		TLLLFIPVLVLTACGTLTGILAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAX				
orf83a		MKTLLXLIPLVLTACGTLTGIPAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL				
10		10	20	30	40	50
		60	70	80	90	100
orf83.pep		YVSVMGDQSGNISGGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS				
15		70	80	90	100	110
orf83a		YVSVMGDQSGNISGGGRYSIDALIRGGYHNNPESATQYSYPAYDTTATTKSDALSSVTTS				
		70	80	90	100	110
		120	130	140	150	160
orf83.pep		TSLLNAPAAAXLTKN SGRK GERSAGLSVNGTGDYRNETLLANPRDVSFLTNIQT V FYLRG				
20		130	140	150	160	170
orf83a		TSLLNAPAAALTKN SGRK GERSAGLSVNGTGDYRNETLLANPRDVSFLTNIQT V FYLRG				
		130	140	150	160	170
		180	190			
orf83.pep		IEVVPXYADTDV FVTVDV				
25		190	200	210	220	230
orf83a		IEVVPPEYADTDV FVTVDV FGT VRSRTELHLYNAETLKAQTKLEYFAVDRDRSKLLIAPK				
		190	200	210	220	230

The complete length ORF83a nucleotide sequence <SEQ ID 315> is:

30	1	ATGAAAACCC	TGCTCNCCT	CATCCCCCTC	GTCCTCACAG	CCTGCGGCAC
	51	ACTGACCGGC	ATACCCGCC	ACGGCGGCGG	CAAACGCTTT	GCCGTCGAAC
	101	AAGAACTCGT	CGCCGCATCG	TCCC GCGCCG	CCGTCAAAGA	AATGGACTTG
	151	TCCGCCCTGA	AAGGACGCAA	AGCCGCCCTT	TACGTCTCCG	TTATGGGCGA
35	201	CCAAGGTTCG	GGCAACATAA	GCGGCGGACG	CTACTCTATC	GACGCACTGA
	251	TACGCGGCGG	CTACCACAAC	AACCCCGAAA	GTGCCACCCA	ATACAGCTAC
	301	CCGCGCTACG	ACACTACCGC	CACCACCAAA	TCCGACGCGC	TCTCCAGCGT
	351	AACCACTTCC	ACATCGCTTT	TGAACGCCCC	CGCCGCCGCC	CTGACGAAAA
	401	ACAGCGGACG	CAAAGGCGAA	CGCTCCGCCG	GACTGTCCGT	CAACGGCACG
40	451	GGCGACTACC	GCAACGAAAC	CCTGCTCGCC	AACCCCGCG	ACGTTTCCTT
	501	CCTGACCAAC	CTCATCCAAA	CCGTCTTCTA	CCTGCGCGGC	ATCGAAGTCG
	551	TACCGCCCGA	ATACGCCGAC	ACCGACGTAT	TCGTAACCGT	CGACGTATTC
	601	GGCACCCTCC	GCAGCCGCAC	CGAACTGCAC	CTCTACAACG	CCGAAACCTT
	651	TAAAGCCCAA	ACCAAGCTCG	AATATTTCGC	CGTTGACCGC	GACAGCCGGA
45	701	AACTGCTGAT	TGCCCCTAAA	ACCGCCGCCT	ACGAATCCCA	ATACCAAGAA
	751	CAATACGCCC	TCTGGATGGG	ACCTTACAGC	GTCGGCAAAA	CCGTCAAAGC
	801	CTCAGACCGC	CTGATGGTCG	ATTTCTCCGA	CATCACCCCC	TACGGCGACA
	851	CAACCGCCCA	AAACCGTCCC	GACTTCAAAC	AAAACAACGG	TAAAAAACCC
	901	GATGTCGGCA	ACGAAGTCAT	CCGCCGCCG	AAAGGAGGAT	AA

This encodes a protein having amino acid sequence <SEQ ID 316>:

50	1	MKTLLXLIPL	VLTACGTLTG	IPAHGGGKRF	AVEQELVAAS	SRAAVKEMDL
	51	SALKGRKAAL	YVSVMGDQGS	GNISGGGRYSI	DALIRGGYHN	NPESATQYSY
	101	PAYDTTATTK	SDALSSVTTS	TSLLNAPAAA	LTKN SGRKGE	RSAGLSVNGT
	151	GDYRNETLLA	NPRDVSFLT	NIQT V FYLRG	IEVVPPEYAD	TDV FVTVDV
	201	GT VRSRTELH	LYNAETLKAQ	TKLEYFAVDR	DSRKLIIAPK	TAAYESQYQE
55	251	QYALWMPYS	VGKTVKASDR	LMVDFSDITP	YGDTTAQNR	DFKQNGKKP
	301	DVGNEVIRRR	KGG*			

ORF83a and ORF83-1 show 98.4% identity in 313 aa overlap:

	10	20	30	40	50	60
orf83a.pep	MKTLLXLIPLVLTACGTLTGIPAHGGGKRF AVEQELVAASSRAAVKEMDLSALKGRKAAL					

25

ORF83 shows 94.9% identity over a 197aa overlap with a predicted ORF (ORF83.ng) from *N.*

gonorrhoeae:

50

The complete length ORF83ng nucleotide sequence <SEQ ID 317> is:

65

-213-

5 651 TAAAGCCCAA ACCAAGCTCG AATATTTTCGC CGTCGACCGC GACAGCCGGA
 701 AACTGCTGAT TGCCCCCTAAA ACCGCCGCCT ACGAATCCCA ATACCAAGAA
 751 CAATACGCCC TCTGGATGGG ACCTTACAGC GTCGGCAAAA CCGTCAAAGC
 801 CTCAGACCGC CTGATGGTCG ATTTCTCCGA CATCACCCCC TACGGCGACA
 851 CAACCGCCCA AAACCGTCCC GACTTCAAAC AAAACAACGG TAAAAACCCC
 901 GATGTCGGCA ACGAAGTCAT CCGCCGCCGC AAAGGAGGAT AA

This encodes a protein having amino acid sequence <SEQ ID 318>:

10 1 MKTLLLLLIPL VLTACGTLTG IPAHHGGGKRF AVEQELVAAS SRAAVKEMDL
 51 SALKGRKAAL YVSVMGDQGS GNISGGYRSI DALIRGGYHN NPDSATRYSY
 101 PAYDTTATTK SDALSGVTTS TSLNAPAAA LTKNNGRKGE RSAGLSVNGT
 151 GDYRNETLLA NPRDVSFLTNI LIQTVFYLRG IEVVPPEYAD TDVFVTVDFV
 201 GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLIAPK TAAYESQYQE
 251 QYALWMGPYS VGKTVKASDR LMVDFSDITP YGDTTAQNRP DFKQNNNGKNP
 301 DVGNEVIRRR KGG*

15 ORF83ng and ORF83-1 show 97.1% identity in 313 aa overlap

		10	20	30	40	50	60
orf83-1.pep		MKTLLLLLIPLVLTACGTLTGIPAHHGGGKRF	AVEQELVAASSRAAVKEMDLSALKGRKAAL				
20 orf83ng		MKTLLLLLIPLVLTACGTLTGIPAHHGGGKRF	AVEQELVAASSRAAVKEMDLSALKGRKAAL				
		10	20	30	40	50	60
		70	80	90	100	110	120
orf83-1.pep		YVSVMGDQGS	GNISGGYRSIDALIRGGYHNPESATQYSYPAYDTTATTKSDALSSVTTS				
25 orf83ng		YVSVMGDQGS	GNISGGYRSIDALIRGGYHNPDSATRYSYPAYDTTATTKSDALSGVTTS				
		70	80	90	100	110	120
		130	140	150	160	170	180
orf83-1.pep		TSLNAPAAALTKNSGRKGRSAGLSVNGTGDYRNETLLANPRDVSFLTNIQTVFYLRG					
30 orf83ng		TSLNAPAAALTKNNGRKGRSAGLSVNGTGDYRNETLLANPRDVSFLTNIQTVFYLRG					
		130	140	150	160	170	180
		190	200	210	220	230	240
orf83-1.pep		IEVVPPEYADTDVFVTVDFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLITPK					
35 orf83ng		IEVVPPEYADTDVFVTVDFGTVRSRTELHLYNAETLKAQTKLEYFAVDRDSRKLITAPK					
		190	200	210	220	230	240
		250	260	270	280	290	300
orf83-1.pep		TAAYESQYQEYALWTGPYKVS	TKASDRLMVDFSDITPYGDTTAQNRPDFKQNNNGKKP				
45 orf83ng		TAAYESQYQEYALWMGPYSVGKTVKASDRLMVDFSDITPYGDTTAQNRPDFKQNNNGKNP					
		250	260	270	280	290	300
		310					
orf83-1.pep		DVGNEVIRRRKGGX					
50 orf83ng		DVGNEVIRRRKGGX					
		310					

Based on this analysis, including the presence of a putative ATP/GTP-binding site motif A (P-loop) in the gonococcal protein (double-underlined) and a putative prokaryotic membrane lipoprotein lipid attachment site (single-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 38

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 319>:

```

1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAACATT
5  51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
101 AAGCCATACG CCGTAAAGTA TTTACGAACA TAAAAGGCTT GAAAATACCG
151 CACACCTACA TAGAAACGGA CGCAAAAAAG CTGCCGAAAT CGACAGATGA
201 GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
251 TCGGGTCTAT TGTCAATTGA GATGAAGCTC AAGACGTATG GCCGGCACGC
10  301 TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
351 ACATCAGGGC ATTGATATAT TTGTTTGTAC TCAAGGTCTT AAGCTTCTAG
401 ATCAAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGCTTCAAAC
451 AAGATGGGTA TGCCTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC
501 CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
15  551 AAGTTTATGA CTGTATGAT TTTAGCGGAA TTTATACCGT AAATAAGGTC
601 AAGCGGTCAA AGTGGTTTTA CACTCTGCCA GTAATAGTAT TGCTGATTCC
651 CGTGTTTGTC GGCCTGTCCT ATAAAATGTT GagCaGTTAC GGAAAAAAC
701 aGGAAGAACC CGCAGCACAA GAATCGGCGG CAACAGAACA GCAGGCAGTA
751 CTTCCGGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
20  801 AGATATGTTT GTTCCGACAT TGTCCGAAaA ACCCGrAAGC AAGCcgaTTT
851 ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
901 GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCaTCAAG GGACGGCATT
951 gaAAGAAGTG ACGGaGTTGA TGTGcgaAgG aCTATGTaAA AAacGGCTTG
1001 CCGTTTAACC CaTACAAAGA AGAAAGCCAA GGGCAGGAAG TTCAGCAAAG
25  1051 CGCGCagCAA CATTCCGACA GGGCGgCAAG TTGCCACATT GGGCGGAAAA
1101 CCGTAGCAGA ACCTAATGTA CGATAATTGG GAAGACGCG GGAACCGTT
1151 TGAAGGAATC GGgCGGGGCG GTGGTCGGAT CGGCAACTG A

```

This corresponds to the amino acid sequence <SEQ ID 320; ORF84>:

```

1  MAEICLITGT PGSGKTLKMW SMMANDEMFK PDEKAIRRVK FTNIKGLKIP
30  51  HTYIETDAKK LPKSTDEQLS AHDMEWIKK PENIGSIVIV DEAQDVWPAR
101 SAGSKIENV QWLNTHRHQG IDIFVLTQGP KLLDQNLRTL VRKHYHIASN
151 KMGMRLLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYX XAEVHTVNKV
201 KRSKWFTLP VIVLLIPVFV GLSYKMLSSY GKKQEEPAAQ ESAATEQQAV
251 LPDKTEGEPV NNGNLADMF VPTLSEKPKS KPIYNGVRQV RTFEYIAGCI
35  301 EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGOEVQQS
351 AQQHSRAQV ATLGKPKXQN LMYDNWEERG KPFEIGGGV VGSAN*

```

Further work revealed the complete nucleotide sequence <SEQ ID 321>:

```

1  ATGGCAGAGA TCTGTTTGAT AACCGGCACG CCCGGTTCAG GGAAACATT
40  51  AAAAATGGTT TCCATGATGG CGAATGATGA AATGTTTAAG CCTGATGAAA
101 ACGGCATACG CCGTAAAGTA TTTACGAACA TAAAAGGCTT GAAAATACCG
151 CACACCTACA TAGAAACGGA CGCAAAAAAG CTGCCGAAAT CGACAGATGA
201 GCAGCTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
251 TCGGGTCTAT TGTCAATTGA GATGAAGCTC AAGACGTATG GCCGGCACGC
45  301 TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
351 ACATCAGGGC ATTGATATAT TTGTTTGTAC TCAAGGTCTT AAGCTTCTAG
401 ATCAAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGCTTCAAAC
451 AAGATGGGTA TGCCTACGCT TTTAGAATGG AAAATATGCG CGGACGATCC
501 CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
55  551 AAGTTTATGA CTGTACGAA TCAGCGGAAG TTCATACCGT AAATAAGGTC
601 AAGCGGTCAA AGTGGTTTTA CACTCTGCCA GTAATAGTAT TGCTGATTCC
651 CGTGTTTGTC GGCCTGTCCT ATAAAATGTT GAGCAGTTAC GGAAAAAAC
701 AGGAAGAACC CGCAGCACAA GAATCGGCGG CAACAGAACA GCAGGCAGTA
751 CTTCCGGATA AAACAGAAGG CGAGCCGGTA AATAACGGCA ACCTTACCGC
801 AGATATGTTT GTTCCGACAT TGTCCGAAA ACCCGAAAGC AAGCCGATTT
55  851 ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTATA
901 GAAGGCGGAA GAACCGGATG CGCCTGCTAT TCGCATCAAG GGACGGCATT
951 GAAAGAAGTG ACGGAGTTGA TGTGCAAGGA CTATGTAAAA AACGGCTTGC
1001 CGTTTAACCC ATACAAAGAA GAAAGCCAAG GGCAGGAAGT TCAGCAAAGC
60  1051 GCGCAGCAAC ATTCCGACAG GGCCAAGTT GCCACATTGG GCGGAAAACC
1101 GTAGCAGAAC CTAATGTACG ATAATTGGGA AGAACGCGGG AAACCGTTTG
1151 AAGGAATCGG CGGGGGCGTG GTCGATCGG CAACTGA

```

This corresponds to the amino acid sequence <SEQ ID 322; ORF84-1>:

1 MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRKV FTNIKGLKIP
 51 HTYIETDAKK LPKSTDEQLS AHDMYEWIKK PENIGSIVIV DEAQDVWPAR
 101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGP KLLDQNLRTL VRKHYHIASN
 151 KMGMRITLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYE SAEVHTVNVK
 5 201 KRSKWFTLTP VIVLLIPVFV GLSYKMLSSY GKKQEEPAQ ESAATEQQAV
 251 LPDKTEGEPV NNGNLTADMV VPTLSEKPES KPIYNGVRQV RTFEYIAGCI
 301 EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGEVQQS
 351 AQQHSRAQV ATLGKPK*QN LMYDNWEERG KPFEGIGGGV VGSAN*

Computer analysis of this amino acid sequence gave the following results:

10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF84 shows 93.9% identity over a 395aa overlap with an ORF (ORF84a) from strain A of *N.*

meningitidis:

15	orf84.pep	MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENKIRRKV FTNIKGLKIP HTYIETDAKK
	orf84a	MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRKV FTNIKGLKIP HTYIETDAKK
20	orf84.pep	LPKSTDEQLSAHDMYEWIKK PENIGSIVIV DEAQDVWPAR SAGSKIPENV QWLNTHRHQG
	orf84a	LPKSTDEQLSAHDMYEWIKK PENIGSIVIV DEAQDVWPAR SAGSKIPENV QWLNTHRHQG
25	orf84.pep	IDIFVLTQGP KLLDQNLRTL VRKHYHIASN KMGMRITLLEW KICADDPVKM ASSAFSSIYT
	orf84a	IDIFVLTQGS KLLDQNLRTL VRKHYHIASN KMGMRITLLEW KICADDPVKM ASSAFSSIYT
30	orf84.pep	LDKKVYDLYXAEVHTVNVK KRSKWFTLTP VIVLLIPVFV GLSYKMLSSY GKKQEEPAQ
	orf84a	LDKKVYDLYEAEVHTVNVK KRSKWFTLTP VIVLLIPVFV GLSYKMLSSY GKKQEEPAQ
35	orf84.pep	ESAATEQQAV LPDKTEGEPV NNGNLTADMV VPTLSEKPES KPIYNGVRQV RTFEYIAGCI
	orf84a	ESAATEHQAV FQDKTEGEPV NNGNLTADMV VPTLSEKPES KPIYNGVRQV RTFEYIAGCV
40	orf84.pep	EGGRTGCACY SHQGTALKEV TELMCKDYVK NGLPFNPYKE ESQGEVQQS AQQHSRAQV
	orf84a	EGGRTGCTCY SHQGTALKEIT KEMCKDYARNGL PFNPYKEESQGRDQQSE QHSDRPOV
45	orf84.pep	ATLGKPKXQNL MYDNWEERGK PFEGIGGGV VGSANX
	orf84a	ATLGKPKWQNL MYDNWQERGK PFEGIGGGV VGSANX

The complete length ORF84a nucleotide sequence <SEQ ID 323> is:

55 1 ATGGCAGAGA TCTGTTTAT AACCGGCACG CCCGGTTCAG GGAAACATT
 51 AAAAATGGTT TCCATGATGG CAAACGATGA AATGTTTAAG CCGGATGAA
 101 ACGGCATACG CCGTAAAGTA TTTACGAACA TCAAAGGCTT GAAGATACCG
 151 CACACCTACA TAGAAACGGA CGCGAAAAAG CTGCCGAAAT CGACAGATGA
 201 GCAGGTTTCG GCGCATGATA TGTACGAATG GATAAAGAAG CCCGAAAATA
 251 TCGGGTCTAT TGTCATTGTA GATGAAGCTC AAGACGTATG GCCGGCACGC
 301 TCGGCAGGTT CAAAAATCCC TGAAAATGTC CAATGGCTGA ATACGCACAG
 351 ACATCAGGGC ATTGATATAT TTGTTTGTAC TCAAGGCTCT AAGCTTCTAG
 401 ATCAAATCT TAGAACGCTT GTACGGAAAC ATTACCACAT CGCTTCAAAC
 451 AAGATGGGTA TCGGTACGCT TTTAGATGG AAAATATGCG CGGACGATCC

-216-

501 CGTAAAAATG GCATCAAGCG CATTCTCCAG TATCTATACA CTGGATAAAA
 551 AAGTTTATGA CTTGTACGAA TCAGCGGAAG TTCATACCGT AAATAAGGTC
 601 AAGCGGTCAA AATGGTTTAA TACTCTGCCA GTAATAATAT TGCTGATTCC
 651 CGTTTTGTGC GGCCTGTCCCT ATAAATGTT AAGTAGTTAT GGAAAAAAC
 701 AGGAAGAACC CGCAGCACA GAATCGGCGG CAACAGAACA TCAGGCAGTA
 751 TTTCAGGATA AAACAGAAGG CGAGCCGGTA AACACGGTA ACCTTACCGC
 801 AGATATGTTT GTTCCGACAT TGTCCGAAAA ACCCGAAAGC AAGCCGATTT
 851 ATAACGGTGT AAGGCAGGTA AGAACCTTTG AATATATAGC AGGCTGTGTA
 901 GAAGGCGGAA GAACCGGATG CACATGCTAT TCGCATCAAG GGACGGCATT
 951 GAAAGAAATT ACAAAGGAAA TGTGCAAGGA TTACGCAAGA AACGGATTGC
 1001 CGTTTAACCC ATATAAGAA GAAAGCCAAG GCGCGGATGT CCAGCAAAGT
 1051 GAGCAGCACC ATTCGGACAG ACCGCAAGTT GCCACGTTGG GCGGAAAGCC
 1101 GTGGCAAAAT CTTATGTATG ATAATGGCA GGAGCGCGGA AAACCGTTTG
 1151 AAGGAATCGG CGGGGGCGTG GTCGGATCGG CAAACTGA

15 This encodes a protein having amino acid sequence <SEQ ID 324>:

1 MAEICLITGT PGSGKTLKMV SMMANDEMFK PDENGIRRV FTKIKGLKIP
 51 HTYIETDAKK LPKSTDEQLS AHDMEYIWK PENIGSIVIV DEAQDVWPAR
 101 SAGSKIPENV QWLNTHRHQG IDIFVLTQGS KLLDQNLRTL VRKHYHIASN
 151 KMGMRITLLEW KICADDPVKM ASSAFSSIYT LDKKVYDLYE SAEVHTVNVK
 201 KRKSWFYTLF VIILLIPVFV GLSYKMLSSY GKQKEEPAAQ ESAATEHQAV
 251 FQDKTEGEPV NNGNLTADMV VPTLSEKPEP KPIYNGVRQV RTFEYIAGCV
 301 EGGRTGCTCY SHQGTALKEI TKEMCKDYAR NGLPFNPYKE ESQGRDVQQS
 351 EQHHSRDPQV ATLGKPKWQN LMYDNWQERG KPFEGIGGGV VGSAN*

ORF84a and ORF84-1 show 95.2% identity in 395 aa overlap:

25	orf84a.pep	10	20	30	40	50	60
	MAEICLITGT	PGSGKTLKMV	SMMANDEMFK	PDENGIRRV	FTNIKGLKIP	HTYIETDAKK	
	orf84-1	MAEICLITGT	PGSGKTLKMV	SMMANDEMFK	PDENGIRRV	FTNIKGLKIP	HTYIETDAKK
30	orf84a.pep	70	80	90	100	110	120
	LPKSTDEQLS	AHDMEYIWK	PENIGSIVIV	DEAQDVWPAR	SAGSKIPENV	QWLNTHRHQG	
	orf84-1	LPKSTDEQLS	AHDMEYIWK	PENIGSIVIV	DEAQDVWPAR	SAGSKIPENV	QWLNTHRHQG
35	orf84a.pep	130	140	150	160	170	180
	IDIFVLTQGS	KLLDQNLRTL	VRKHYHIASN	KMGMRITLLEW	KICADDPVKM	ASSAFSSIYT	
	orf84-1	IDIFVLTQGS	KLLDQNLRTL	VRKHYHIASN	KMGMRITLLEW	KICADDPVKM	ASSAFSSIYT
40	orf84a.pep	190	200	210	220	230	240
	LDKKVYDLYE	SAEVHTVNVK	KRKSIFYTLF	VIILLIPVFV	GLSYKMLSSY	GKQKEEPAAQ	
	orf84-1	LDKKVYDLYE	SAEVHTVNVK	KRKSIFYTLF	VIILLIPVFV	GLSYKMLSSY	GKQKEEPAAQ
45	orf84a.pep	250	260	270	280	290	300
	ESAATEHQAV	FQDKTEGEPV	NNGNLTADMV	FVPTLSEKPEP	KPIYNGVRQV	RTFEYIAGCV	
	orf84-1	ESAATEHQAV	FQDKTEGEPV	NNGNLTADMV	FVPTLSEKPEP	KPIYNGVRQV	RTFEYIAGCV
50	orf84a.pep	310	320	330	340	350	360
	EGGRTGCTCY	SHQGTALKEI	TKEMCKDYAR	NGLPFNPYKE	ESQGRDVQQS	EQHHSRDPQV	
	orf84-1	EGGRTGCTCY	SHQGTALKEI	TKEMCKDYAR	NGLPFNPYKE	ESQGRDVQQS	EQHHSRDPQV
55	orf84a.pep	370	380	390			
	ATLGKPKWQN	LMYDNWQERG	KPFEGIGGGV	VGSANX			
	orf84-1	ATLGKPKWQN	LMYDNWQERG	KPFEGIGGGV	VGSANX		
60	orf84a.pep	370	380	390			
	ATLGKPKWQN	LMYDNWQERG	KPFEGIGGGV	VGSANX			
65	orf84-1	ATLGKPKWQN	LMYDNWQERG	KPFEGIGGGV	VGSANX		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF84 shows 94.2% identity over a 395aa overlap with a predicted ORF (ORF84.ng) from *N.*

gonorrhoeae:

5	orf84.pep	MAEICLITGTPGSGKTLKVM	SMMANDEMFKPDEKAIRRKVFTNIKGLKIPHTYIETDAKK	60
	orf84ng	MAEICLITGTPGSGKTLKVM	SMMANDEMFKPDENGVRKVFTNIKGLKIPHTIETDAKK	60
10	orf84.pep	LPKSTDEQLSAHDMYEWIKK	PENIGSIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG	120
	orf84ng	LPKSTDEQLSAHDMYEWIKK	PENVGAIIVIVDEAQDVWPARSAGSKIPENVQWLNTHRHQG	120
15	orf84.pep	IDIFVLTQGP	KLDDQNLRITLVKRHYHIASNKMGMRTLLEWKICADDPVKMASSAFSSIYT	180
	orf84ng	IDIFVLTQGP	KLDDQNLRITLVKRHYHIAANKMGLRITLLEWKVCADDPVKMASSAFSSIYT	180
20	orf84.pep	LDKKVYDLYXAEVHTV	NKVKRSKWFTLPVIVLLIPVFGLSYKMLSSYGKKQEEPAAQ	240
	orf84ng	LDKKVYDLYSAEIH	TVNKVKRSKWFYALPVIILLIPLFVFGLSYKMLGSGYKKQEEPAAQ	240
25	orf84.pep	ESAATEQQAVLPDKTEGE	SVNNGNLTADMVPTLSEKPKSPIYNGVRQVRTFEYIAGCI	300
	orf84ng	ESAATEQQAVLPDKTEGE	SVNNGNLTADMVPTLPEKPESKPIYNGVRQVRTFEYIAGCI	300
30	orf84.pep	EGGRTGCACYSHQGTALKEV	TELMCKDYVKNGLPFNPYKEESQSQEVQQAQHSRAQV	360
	orf84ng	EGGRTGCTCYSHQGTALKEV	TELMCKDYVKNGLPFNPYKEESQSQEVQQAQHSRAQV	360
35	orf84.pep	ATLGGKPKQNL	MYDNWEERGKPFEGIGGGVVG SAN 395	
	orf84ng	ATLGGKPKQNL	MYDNWEERGKPFEGIGGGVVG SAN 395	

The complete length ORF84ng nucleotide sequence <SEQ ID 325> is:

1	ATGGCAGAAA	TCTGTTTGAT	AACCGGCACG	CCCGGTTTACG	GGAAAAACATT
51	AAAAATGGTT	TCCATGATGG	CAAACGATGA	AATGTTTAAG	CCAGATGAAA
101	ACGGCGTACG	CCGTAAAGTA	TTTACGAACA	TCAAAGGTTT	GAAGATACCG
151	CACACCCACA	TAGAAACAGA	CGCAAAGAAG	CTGCCGAAAT	CAACCGATGA
201	ACAGCTTTCG	GCGCATGATA	TGTATGAATG	GATCAAGAAG	CCTGAAAacg
251	tcggcgCAAT	CGTTATTGTC	GATGAGGCGC	AAGACSTATG	GCCCGCACGC
301	TccgCAGGTT	CGAAAATCCC	CGAAAACGTC	CAATGGCTGA	ACACACACAG
351	GCATCAGGGC	ATAGATATAT	TGTATTGAC	ACAAGTCCT	AACTCTTAG
401	ATCAGAACTT	GCGAACATTG	GTTAAAAGAC	ATTACCACAT	TGCGGCCAAC
451	AAAATGGGTT	TGCGTACCCT	GCTTGAATGG	AAAGTATGCG	CGGATGACCC
501	GGTAAAAATG	GCATCAAGTG	CATTTTCCAG	TATCTACACA	CTGGATAAAA
551	AAGTTTATGA	CTTGACGAA	TCCGCAGAAA	TTCACACGGT	AAACAAAGTC
601	AAGCGTTCAA	AATGGTTTTA	TGCATTGCC	GTCATCATAT	TATTGATTCC
651	GCTATTTGTC	GGTTTGTCTT	ACAAAATGTT	GGGCAGTTAC	GGAAAAAAAC
701	AGGAAGAACC	CGCAGCACAA	GAATCGGCGG	CAACAGAACA	GCAGGCAGTA
751	CTTCCGGATA	AACAGAAGG	AGAATCGGTG	AATAACGGAA	ACCTTACGGC
801	AGATATGTTT	GTTCCGACAT	TGCCCCGAAA	ACCCGAAAGC	AAGCCGATTT
851	ATAACGGTGT	AAGGCAGGTA	AGGACCTTTG	AATATATAGC	AGGCTGTATA
901	GAAGGCGGAA	GAACCGGATG	CACCTGCTAT	TCGCATCAAG	GGACGGCATT
951	GAAAGAAGTG	ACGGAGTTGA	TGTGCAAGGA	CTATGTAAAA	AACGGCTTGC
1001	CGTTTAACCC	ATACAAAGAA	GAAAGCCAAG	GGCAGGAAGT	TCAGCAAAGC
1051	GCGCAGCAAC	ATTCGGACAG	GGCGCAAGTT	GCCACCTTGG	GCGGAAAACC
1101	GCAGCAGAAC	CTAATGTACG	ACAATTGGGA	AGAACGCGGG	AAACCGTTTG
1151	AAGGAATCGG	CGGGGCGGTG	GTCGGATCGG	CAACTGA	

This encodes a protein having amino acid sequence <SEQ ID 326>:

1	MAEICLITGT	PGSGKTLKMV	SMMANDEMFK	PDENGVRKV	FTNIKGLKIP
51	HTHIETDAKK	LPKSTDEQLS	AHDMYEWIKK	PENVGAIVIV	DEAQDVWPAR
101	SAGSKIPENV	QWLNTHRHQG	IDIFVLTQGP	KLDDQNLRITL	VKRHYHIAAN
151	KMGLRITLLEW	KVCADDPVKM	ASSAFSSIYT	LDKKVYDLYE	SAEIHVNKV
201	KRSKWIFYALP	VIILLIPLFV	GLSYKMLGSY	GKKQEEPAAQ	ESAATEQQAV
251	LPDKTEGESV	NNGNLTADMV	VPTLPEKPES	KPIYNGVRQV	RTFEYIAGCI
301	EGGRTGCTCY	SHQGTALKEV	TELMCKDYVK	NGLPFNPYKE	ESQSQEVQQA
351	AQQHSRAQV	ATLGGKPKQN	LYDNWEERG	KPFEGIGGGV	VGSAN*

ORF84ng and ORF84-1 show 95.4% identity in 395 aa overlap:

		10	20	30	40	50	60
	orf84-1.pep	MAEICLITGT	PGSGKTLK	MVSMANDE	MFKPDENG	IRRKVFTN	IKGLKIPHTYIETDAKK
5	orf84ng	MAEICLITGT	PGSGKTLK	MVSMANDE	MFKPDENG	VRRKVFTN	IKGLKIPHTHIETDAKK
		10	20	30	40	50	60
	orf84-1.pep	LPKSTDEQL	SAHDMYEW	IKKPENIG	SIVIVDEA	QDVWPARS	SAGSKIPENVQWLNTHRHQG
10	orf84ng	LPKSTDEQL	SAHDMYEW	IKKPENVG	AIIVDEA	QDVWPARS	SAGSKIPENVQWLNTHRHQG
		70	80	90	100	110	120
	orf84-1.pep	IDIFVLTQ	GPKLLDQ	NLRLTVR	KHYHIAS	NKMGMR	LLEWKICADDPVKMASSAFSSIYT
15	orf84ng	IDIFVLTQ	GPKLLDQ	NLRLTVR	KHYHIA	ANKMGL	RLLLEWKVCADDPVKMASSAFSSIYT
		130	140	150	160	170	180
	orf84-1.pep	LDKKVYD	LYESA	AEVHTV	NKVKRS	KWFYTL	LPVIVLLIPVFGLSYKMLSSYGKKQEPAAG
20	orf84ng	LDKKVYD	LYESA	EIHVTN	KVKRS	KWFYAL	PVILLIPVFGLSYKMLSSYGKKQEPAAG
		190	200	210	220	230	240
	orf84-1.pep	ESAATEQ	QAVLPD	KTEGE	PVNNGL	TADMVPT	LSEKPEKPIYNGVRQVRTFEYIAGCI
25	orf84ng	ESAATEQ	QAVLPD	KTEGE	SVNNGL	TADMVPT	LPEKPEKPIYNGVRQVRTFEYIAGCI
		250	260	270	280	290	300
	orf84-1.pep	EGGRTG	CACYS	HQGTAL	KEVTE	LMCKDY	VKNGLPFNPKYKEESQGGQEVQSSAQQHS
30	orf84ng	EGGRTG	CCTCY	SHQGT	ALKEV	TELMCK	DYVKNGLPFNPKYKEESQGGQEVQSSAQQHS
		310	320	330	340	350	360
	orf84-1.pep	ATLGGK	PXQNL	MYDNW	EERKGP	FEFEG	IGGGVVGGSANX
35	orf84ng	ATLGGK	PQNL	MYDNW	EERKGP	FEFEG	IGGGVVGGSANX
		370	380	390			

Based on this analysis, including the presence of a putative transmembrane domain (single-underlined) in the gonococcal protein, and a putative ATP/GTP-binding site motif A (P-loop, double-underlined), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 39

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 327>:

	1	GTGGTTTCC	TGAATGCCGA	CAACGGGATA	TTGGTTCAGG	ACTTGCCTTT
50	51	TGAAGTCAA	CTGAAAAAAT	TCCATATCGA	TTTTTACAAT	ACGGGTATGC
	101	CGCGTGATT	CGCCAGCGAT	ATTGAAGTGA	CGGACAAGGC	AACCGGTGAG
	151	AAACTCGAGC	GCACCATCCG	CGTGAACCAT	CCTTTGACCT	TGCACGGCAT
	201	CACGATTTAT	CAGGCGAGTT	TTGCCGACGG	CGGTTCCGGAT	TTGACATTCA
	251	AGCGGTGGA	TTTGGGTGAT	GCTTCGCGCG	AGCCTGTCGT	GTTGAAGGCA
55	301	ACATCCATAC	ACCACTTTCC	GTTGGAAATT	GGCAAACACA	AATATCGTCT
	351	TGAGTTCGAT	CAGTTCACCT	CTATGAATGT	GGAGGACATG	AGCGAGGGCG
	401	CGGAACGGGA	AAAAAGCCTG	AAATCCACGC	TGCCCGATGT	CCGCGCCGTT
	451	ACTCAGGAAG	GTCACAAATA	CACCAAT...TACCG	
	501	TATCCGTGAT	GCGCCAGGCC	AGGCGGTCGA	ATATAAAAC	TATATGCTGC
60	551	CGGTTTTCGA	GGAACAGGAT	TATTTTTCGA	TTACCGGCAC	GCGCAGCGC.

5
10
1101

```

601 TTGCAGCAGC AATACCGCTG GCTGCGTATC CCCTTGGACA AGCAGTTGAA
651 AGCGGACACC TTTATGGCAT TGCCTGAGTT TTTGAAAGAT GGGGAAGGGC
701 GCAAACGTCT .GTTGCCGAC GCAACCAAAG GCGCACCTGC CGAAATCCGC
751 GAACAATTCA TGCTGGCTGC GGAACAACAC CTGAACATCT TTGCACAAAA
801 AGGCTATTG GATTGGACG AATTATTAC GTCCAATATC CCGAAAGAGC
851 AGCAGGATAA GATGCAGGGC TATTTCTACG AAATGCTTTA CGGCGTGATG
901 AACGCTGCTT TGGATGAAAC CAT.ACCCGG TACGGCTTGC CCGAATGGCA
951 GCAGGATGAA GCGCGGAATC GTTTCCTGCT GCACAGTATG GATGCGTACA
1001 CGGGTTTGAC CGAATATCCC GCGCCTATGC TGCTGCAACT TGATGGGTTT
1051 TCCGAGGTGC GTTCGTCGGG TTTGCAGATG ACCCGTTCCC C.GTCCCGT
1101 TTTGGTCTAT CTC...
```

This corresponds to the amino acid sequence <SEQ ID 328; ORF88>:

15
20

```

1 MVFLNADNGI LVQDLPEFEV LKKFHIDFYN TGMPRDFASD IEVTDKATGE
51 KLERTIRVNH PLTLHGITIY QASFADGGSD LTFKAWNGLD ASREPVLKAA
101 TSIHQFPLEI GKHKYRLEFD QFTSMNVEDM SEGAEREKSL KSTLPDVRVAV
151 TQEGHKYTNX XXXXYRIRD APQAVEYKN YMLPVLQEQD YFWITGTRSX
201 LQQQYRWLRI PLDKQLKADT FMALEFLKLD GEGRRKXVAD ATKGAPAEIR
251 EQFMLAAENT LNIFAQKGYL GLDEFITSNI PKEQQDKMQG YFYEMLYGVM
301 NAALDETSTR YGLPEWQODE ARNRFLHSM DAYTGLTEYP APMLLQLDGF
351 SEVRSSGLQM TRSXGPELLVY L...
```

Further work revealed the complete nucleotide sequence <SEQ ID 329>:

25
30
35
40
45
50
55
60

```

1 ATGAGTAAAT CCCGTAGATC TCCCCCACTT CTTTCCCGTC CGTGGTTCGC
51 TTTTTCAGC TCCATGCGCT TTGCAGTCGC TTTGCTCAGT CTGCTGGGTA
101 TTGCATCGGT TATCGGTACG GTGTTGCAGC AAAACCAGCC GCAGACGGAT
151 TATTTGGTCA AATTCGGATC GTTTTGGGCG CAGATTTTTC GTTTTCTGGG
201 ACTGTATGAC GTCTATGCTT CGGCATGGTT TGTCGTTATC ATGATGTTTT
251 TGGTGGTTTC TACCACTTTG TGCCTGATTC GCAATGTGCC GCCGTTCTGG
301 CGCGAAATGA AGTCTTTTCG GGAAAAGGTT AAAGAAAAAT CTCTGGCGGC
351 GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATTGCGCCC GAGGTGGCCA
401 AACGTATCTT GGAAGTACAA GGTTTTCAGG GAAAAACCAT TAACCGTGAA
451 GACGGGTCGG TTCTGATTGC CGCCAAAAAA GGCACAAATGA ACAAATGGGG
501 CTATATCTTT GCCCATGTTG CTTTGATTGT CATTGCGCTG GGCGGGTTGA
551 TAGACAGTAA CCTGCTGTTG AAAGTGGGTA TGCTGACCCG TCGGATTGTT
601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT
651 GGGTGCCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
701 AGAGTGCCTG TGTTGGTTTC CTGAATGCCG ACAACGGGAT ATTGGTTCAG
751 GACTTGCTCT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTACAAA
801 TACGGGTATG CCGCGTGATT TCGCCAGCGA TATTGAAGTG ACGGACAAGG
851 CAACCGGTGA GAAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC
901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCCGACG GCGETTCGGA
951 TTTGACATTC AAGGCGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTGC
1001 TGTTGAAGGC AACATCCATA CACCAAGTTC CGTTGGAAAT TGGCAAAACAC
1051 AAATATCGTC TTGAGTTTCA TCAGTTCACT TCTATGAATG TGGAGGACAT
1101 GAGCGAGGGC GCGGAACGGG AAAAAAGCCT GAAATCCACG CTGAACGATG
1151 TCCGCGCCGT TACTCAGGAA GSTAAAAAAT ACACCAATAT CGGCCCTTCC
1201 ATTGTTTACC GTATCCGTGA TCGCGCAGGG CAGGCGGTTC AATATAAAAA
1251 CTATATGCTG CCGGTTTTTC AGGAACAGGA TTATTTTGG ATTACGGGCA
1301 CGCGCAGCGG CTTGCAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC
1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAGA
1401 TGGGGAAGGG CGCAAACGTC TGTTTGCCGA CGCAACCAAA GCGCGACCTG
1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CCGAAAACAC GCTGAACATC
1501 TTTGCACAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAATAT
1551 CCCGAAAGAG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT
1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
1651 CCCGAATGGC AGCAGGATGA AGCGCGGAAT CGTTTCCTGC TGCACAGTAT
1701 GGATGCGTAC ACGGGTTTGA CCGAATATCC CGCGCCTATG CTGCTGCAAC
1751 TTGATGGGTT TTCCGAGGTG CGTTCGTCGG GTTTGCAGAT GACCCGTTCC
1801 CCGGGTGCGC TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGTAC
1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GCGGTGGGTA TTGTTTTCAG
1901 ACGGCAAAAT CCGTTTTCGC ATGTCTTCGG CCCGACGCGA ACGGGATTG
1951 CAGAAGGAAT TTCCAAAACA CGTCGAGAGT CTGCAACGGC TCGGCAAGGA
2001 CTTGAATCAT GACTGA
```

This corresponds to the amino acid sequence <SEQ ID 330; ORF88-1>:

65

```

1 MSKSRSPPL LSRPFAFFS SMRFAVALLS LLGIASVIGT VLQONQPTD
51 YLVKFGSFWA QIFGLGLYD VYASAEVVI MMFLVSTSL CLIRNVPPFW
```

5
10
15
20
25
30
35
40
45
50
55
60
65

101 REMKSFREKV KEKSLAAMRH SLLDVKIAP EVAKRYLEVQ GFQGKTINRE
151 DGSVLIAAKK GTMKNWGYIF AHVALIVICL GGLIDSNLL KLGMLTGRIV
201 PDNQAVYAKD FKPEILGAS NLSFRGNVNI SEGQSADVVF LNADNGILVQ
251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT
301 LHGITYQAS FADGGSDLTF KAWNLDASR EPVVLKATSI HQFPLEIGKH
351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS
401 IVYRIRDAAG QAVEYKNYML PVLQEQDYFW ITGTRSGLOQ QYRWLRIPLD
451 KQLKADTFMA LREFLDGEG RKRLVADATK GAPAEIREQF MAAENTLNI
501 FAQKGYLGLD EFITSNIPKE QQDKMQGYFY EMLYGMNAA LDETIRRYGL
551 PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQDGFSEV RSSGLQMTSR
601 PGALLVYLG S VLLVLGTVLM FYVREKRAWV LFSGDKIRFA MSSARSERDL
651 QKEFPKHVES LQRLGKDLNH D*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

15 ORF88 shows 95.7% identity over a 371aa overlap with an ORF (ORF88a) from strain A of *N. meningitidis*:

20	orf88a	AKDFKPESILGASNLSFRGNVNI SEGQSADVVF LNADNGILVQDLPFEVKLKKFHIDFYN
		210 220 230 240 250 260
25	orf88a	TGMPRDFASDIEVTDKATGEKLE RTIRVNHPLTLHGITYQASFADGGSDLTFKAWNLD
		270 280 290 300 310 320
30	orf88a	ASREPVLKATSIHQFPLEIGKHKYRLEFDQFTSMNVEDMSEGAEREKSLKSTLPDVRV
		330 340 350 360 370 380
35	orf88a	TQEGHKYTNXXXXXYRIRDAAGQAVEYKNYMLPVLQEQDYFWITGTRSLQOQQYRWLR
		390 400 410 420 430 440
40	orf88a	PLDKQLKADTFMALREFLDGEGRKRLVADATKGAPAEIREQFMAAENTLNI FAQKGYL
		450 460 470 480 490 500
45	orf88a	GLDEFITSNIPKEQQDKMQGYFY EMLYGMNAAALDETXYTRYGLPEWQQDEARNRFLHSM
		510 520 530 540 550 560
50	orf88a	DAYTGLTEYPAPM LLQDGFSEVRSSGLQMTSPGALLVYLGSVLLVLGTVLMFYVREKR
		570 580 590 600 610 620
55	orf88a	AWLFSGDKIRFAMSSARSERDLQKEFPKHVESLQRLGKDLNHDX
		630 640 650 660 670

The complete length ORF88a nucleotide sequence <SEQ ID 331> is:

1 ATGAGTAAAT CCCGTAGATC TCCCCACTT CTTTCCCGTC CGTGGTTCCG
51 TTTTTCAGC TCCATGCGCT TTGCGGTGCG TTTGCTCAGT CTGCTGGGTA
101 TTGCATCGGT TATCGGTACG GTGTTGCAGC AAAACCAGCC GCAGACGGAT

151 TATTGGTCA AATTCGGATC GTTTTGGGCG CAGATTTTGT GTTTCTGGG
 201 ACTGTATGAC GTCTATGCTT CGGCATGGTT TGTGCTTATC ATGATGTTTT
 251 TGGTGGTTTC TACCACTTTG TGCCTGATTG GCAATGTGCC GCCGTCTCGG
 301 CGCGAAATGA AGTCTTTTCG GGAAAAGGTT AAAGAAAAAT CTCTGGCGGC
 5 351 GATGCGCCAT TCTTCGCTGT TGGATGTAAG AATTGCGCCC GAGGTGCGCA
 401 AACGTTATCT GGAAGTACAA GGTTCCTCAG GAAAAACCAT TAACCGTGAA
 451 GACGGGTCGG TTCTGATTGC CGCCAAAAAA GGCACAATGA ACAATGGGG
 501 CTATATCTTT GCCCATGTTG CTTTGATTGT CATTGCGCTG GCGGGTTGA
 551 TAGACAGTAA CCTGCTGTTG AAACCTGGGTG TGCTGACCGG TCGGATTGTT
 10 601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCC AAAGTATTTT
 651 GGGTGCCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
 701 AGAGTGCCTG TGTGGTTTTC CTGAATGCCG ACAACGGGAT ATTGGTTCAG
 751 GACTTGCCCTT TTGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTACAA
 801 TACGGGTATG CCGCGCGATT TTGCCAGTGA TATTGAAGTA ACGGATAAGG
 15 851 CAACCGGTGA GAACTCTGAG CGCACCATCC GCGTGAACCA TCCTTTGACC
 901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCCGACG GCGGTTGCGA
 951 TTTGACATTC AAGGCGTGGA ATTTGGGTGA TGCTTCGCGC GAGCCTGTCTG
 1001 TGTGAAGGC AACATCCATA CACCAAGTTT CGTTGGAAAT TGGCAAAACAC
 1051 AAATATCGTC TTGAGTTCTG TCAGTTTACT TCTATGAATG TGGAGGACAT
 20 1101 GAGCGAGGGC GCGGAACGGG AAAAAAGCCT GAAATCCACG CTGAACGATG
 1151 TCCGCGCCGT TACTCAGGAA GGTAAGAAAT ACACCAATAT CGGCCCTTCC
 1201 ATTGTTTACC GTATCCGTGA TCGGCGAGGG CAGGCGGTCTG AATATAAAAA
 1251 CTATATGCTG CCGGTTTTGC AGGAACAGGA TTATTTTGG ATTACCGCA
 25 1301 CGCGCAGCGG CTTCGAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGAC
 1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTGAAAGA
 1401 TGGGGAAGGG CGCAAAACGTC TGGTTGCCGA CGCAACCAAA GCGCACCTG
 1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAACAC GCTGAACATC
 1501 TTTGCACAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCCAATAT
 30 1551 CCCGAAAGAG CAGCAGGATA AGATGCAGGG CTATTTCTAC GAAATGCTTT
 1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCGG GTACGGCTTG
 1651 CCCGAATGCG AGCAGGATGA AGCGCGGAAT CGTTTCCTGC TGCACAGTAT
 1701 GGATGCGTAC ACGGGTTTGA CCGAATATCC CGCGCCTATG CTGCTGCAAC
 1751 TTGATGGGTT TCCGAGGTG CGTTCGTCGG GTTTGCAGAT GACCCGTTCC
 1801 CCGGGTGCCG TTTTGGTCTA TCTCGGCTCG GTGCTGTTGG TATTGGGTAC
 35 1851 GGTATTGATG TTTTATGTGC GCGAAAAACG GCGGTGGGTA TTGTTTTCAG
 1901 ACGGCAAAAT CCGTTTTGCC ATGTCTTCGG CCCGCAGCGA ACGGGATTG
 1951 CAGAAGGAAT TTCCAAAACA CGTCGAGAGT CTGCAACGGC TCGGCAAGGA
 2001 CTTGAATCAT GACTGA

This encodes a protein having amino acid sequence <SEQ ID 332>:

40 1 MSKSRRSPPL LSRPWFAPFS SMRFAVALLS LLGIASVIGT VLQONQPQTD
 51 YLVKFGSFWA QIFGFLGLYD VYASAWFVVI MMFLVSTSL CLIRNVPPFW
 101 REMKSFREKV KEKSLAAMRH SLLDVKIAP EVAKRYLEVQ GFQKNTINRE
 151 DGSVLIAAKK GTMNKGYIF AHVALIVICL GGLIDSNLLL KLGLMTGRIV
 201 PDNQAVYAKD FKPEILGAS NLSFRGNVNI SEGQSAADVVF LNAADNGILVQ
 45 251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNEPLT
 301 LHGITYQAS FADGGSDLTF KAWNLDASR EPVVLKATSI HQFPLEIGKH
 351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKTYNIGPS
 401 IVYRIRDAAG QAVEYKNYML PVLQEODYFW ITGTRSGLQQ QYRWLRIPLD
 451 KQLKADTFMA LREFLDGEG RKRLVADATK GAPAEIREQF MLAAENTLNI
 50 501 FAQKGYLGLD EFITSNIPKE QQDKMQGYFY EMLYGVMNAA LDETIRRYGL
 551 PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTSRS
 601 PGALLVYLGS VLLVLGTVLM FYVREKRAWV LFSQDKIRFA MSSARSERDL
 651 QKEFPKHVES LQRLGKDLNH D*

ORF88a and ORF88-1 100.0% identity in 671 aa overlap:

55 orf88a.pep MSKSRRSPPLLSRPWFAPFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA 60
 orf88-1 MSKSRRSPPLLSRPWFAPFSSMRFAVALLSLLGIASVIGTVLQONQPQTDYLVKFGSFWA 60
 60 orf88a.pep QIFGFLGLYDVYASAWFVVIIMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120
 orf88-1 QIFGFLGLYDVYASAWFVVIIMFLVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120
 orf88a.pep SLLDVKIAPEVAKRYLEVQGFQKNTINREDGSVLIAAKKGTMTNKGWYIFAHVALIVICL 180
 65 orf88-1 SLLDVKIAPEVAKRYLEVQGFQKNTINREDGSVLIAAKKGTMTNKGWYIFAHVALIVICL 180
 orf88a.pep GGLIDSNLLLKLGLMTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSAADVVF 240

	orf88-1	 GGLIDSNLLKLGMLTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNI SEGQSADVVF	240
5	orf88a.pep	LNADNGILVQDL PFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT	300
	orf88-1	 LNADNGILVQDL PFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT	300
10	orf88a.pep	LHGITIYQAS FADGGS DLT FKA WN LGDAS REP VV LKATS IHQF PLE IGKHKYRLEFDQFT	360
	orf88-1	 LHGITIYQAS FADGGS DLT FKA WN LGDAS REP VV LKATS IHQF PLE IGKHKYRLEFDQFT	360
15	orf88a.pep	SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
	orf88-1	 SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKNYML	420
	orf88a.pep	PVLQE QDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
	orf88-1	 PVLQE QDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK	480
20	orf88a.pep	GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
	orf88-1	 GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA	540
25	orf88a.pep	LDETIRRYGLPEWQQDEARNRFLHSM DAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
	orf88-1	 LDETIRRYGLPEWQQDEARNRFLHSM DAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS	600
	orf88a.pep	PGALLVYLGSVLLVLGTVMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
30	orf88-1	 PGALLVYLGSVLLVLGTVMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES	660
	orf88a.pep	LQRLGKDLNHD	672
	orf88-1	 LQRLGKDLNHD	672
35			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF88 shows 93.8% identity over a 371aa overlap with a predicted ORF (ORF88.ng) from *N.*

gonorrhoeae:

40	orf88.pep	MVFLNADNGILVQDL PFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNH	60
	orf88ng	 MVFLNADNGMLVQDL PFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNH	60
45	orf88.pep	PLTLHGITIYQAS FADGGS DLT FKA WN LGDAS REP VV LKATS IHQF PLE IGKHKYRLEFD	120
	orf88ng	 PLTLHGITIYQAS FADGGS DLT FKA WN LGDAS REP VV LKATS IHQF PLE IGKHKYRLEFD	120
50	orf88.pep	QFTSMNVEDMSEGAEREKSLKSTLPDVRVAVTQEGHKYTNXXXXXXYRIRDAAGQAVEYKN	180
	orf88ng	 QFTSMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKYTNIGPSIVYRIRDAAGQAVEYKN	180
55	orf88.pep	YMLPVLQE QDYFWITGTRSX LQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRXVAD	240
	orf88ng	 YMLPILQDKDYFWLTGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVAD	240
60	orf88.pep	ATKGAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVM	300
	orf88ng	 ATKDAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKGQQDKMQGYFYEMLYGVM	300
65	orf88.pep	NAALDETXYTRYGLPEWQQDEARNRFLHSM DAYTGLTEYPAPMLLQLDGFSEVRSSGLQM	360
	orf88ng	 NAALDETIRRYGLPEWQQDEARNRFLHSM DAYTGLTEYPAPMLLQLDGFSEVRSSGLQM	360
	orf88.pep	TRSXGPLL VYL	371
	orf88ng	 TRSPGALLVYLGSVLLVLGTVMFYVPPKRAWVLF SNXKIRFAMSSARSERDLQKEFPKH	420

An ORF88ng nucleotide sequence <SEQ ID 333> was predicted to encode a protein having amino acid sequence <SEQ ID 334>:

```

1  MVFLNADNGM LVQDLPFEVK LKKFHIDFYN TGMPRDFASD IEVTDKATGE
5  51  KLERTIRVNH PLTLHGITYI QASFADGGSD LTFKAWNLRD ASREPVVLKA
101 TSIHQFPLEI GKHKYRLEFD QFTSMNVEDM SEGAEREKSL KSTLNDVRAV
151 TQEGKKYTN I GPSIVYRIR AAGQAVEYKN YMLPILQDKD YFWLTGTRSG
201 LQQYRWLRI PLDKQLKADT FMALREFLKD GEGRKRLVAD ATKDAPAEIR
251 EQFMLAAENT LNIFAQKGYL GLDEFITSNI PKGQQDKMQG YFYEMLYGVM
301 NAALDETIRR YGLPEWQQDE ARNRFLHSM DAYTGLTEYP APMLLQLDGF
10 351 SEVRSSGLQM TRSPGALLVY LGSVLLVLGT VFMFYVPKKR AWWLFSNKKI
401 RFAMSSARSE RDLQKEFPKH VESLQRLGKD LNHD*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 335>:

```

1  ATGAGTAAAT CCCGTATATC TCCCACACTT CTTTCCCGTC CGTGTTTCGC
15 51  TTTTTCAGC TCCATGCGCT TTGCGGTCGC TTTGCTCAGT CTGCTGGGTA
101 TTGCATCGGT TATCGGCACG GTGTACAGC AAAACACAGC GCAGACGGAT
151 TATTGGTCA AATTCGGACC GTTTGGACT CGGATTTTGG ATTTTGTGGG
201 TTTGTATGAT GTCTATGCTT CGGCATGGTT TGTCGTTATC ATGATGTTTC
251 TGGTGGTTTC TACCAAGTTG TGTTAATCC GTAACGTTCC GCCGTTTGG
301 CGCGAAATGA AGTCTTCCG GGAAGAGTT AAAGAAAAAT CTCTGGCGGC
20 351 GATGCGCCAT TCTTCGCTGT TGGATGTAAA AATTGCCCCC GAAGTTGCCA
401 AACGTTATCT GGAGGTGCGG GGTTCAGG GAAAAACCGT CAGCCGTGAG
451 GACGGGTCGG TTCTGATTGC CGCCAAAAAA GGCACaatga acaaATGGGG
501 CTATATCTTT GCcgaagtag ctTTGATTGT CATTGCTGCTG GCGGGTTGA
551 TAGACAGTAA CCTGCTGCTG AAGCTGGGTA TGCTGGCCCG TCGGATTGTT
25 601 CCGGACAATC AGGCGGTTTA TGCCAAGGAT TTCAAGCCCG AAAGTATTTT
651 GGGTGCCTCC AATCTCTCAT TTAGGGGCAA CGTCAATATT TCCGAGGGGC
701 AAAGTGCCTG TGTGGTTTTC CTGAATGCCG ACAACGGGAT GTTGGTTTTCAG
751 GACTTGCTTT TGAAGTCAA ACTGAAAAAA TTCCATATCG ATTTTACAA
801 TACGGGTATG CCGCGGATT TTGCCAGCGA TATTGAAGTA ACGGACAAGG
30 851 CAACCGGTGA GAAACTCGAG CGCACCATCC GCGTGAACCA TCCTTTGACC
901 TTGCACGGCA TCACGATTTA TCAGGCGAGT TTTGCCGACG GCGGTTCGGA
951 TTTGACATTC AAGGCGTGGA ATTTGAGGGA TGCTTCGCGC GAACCTGTCTG
1001 TGTTGAAGGC AACCTCCATA CACCAGTTTC CGTTGGAAAT CGGCAAAACAC
1051 AAATATCGTC TTGAGTTCGA TCAGTTCACT TCTATGAATG TGGAGGACAT
35 1101 GAGCGAGGCT GCGGAACGGG AAAAAAGCCT GAAATCCACT CTGAACGATG
1151 TCCGCGCCGT TACTCAGGAA GGTAATAAAT ACACCAATAT CGGCCCTTCC
1201 ATCGTGTACC GCATCCGTGA TGcggCAGGG CAGGCGGTCG AATATAAAAA
1251 CTATATGCTG CCGATTTTGC AGGACAAAGA TTATTTTGG CTGACCGGCA
1301 CGCGCAGCGG CTTGCAGCAG CAATACCGCT GGCTGCGTAT CCCCTTGGAC
40 1351 AAGCAGTTGA AAGCGGACAC CTTTATGGCA TTGCGTGAGT TTTTGAAGA
1401 TGGGGAAGGG CGCAAACGTC TGGTTGCCGA CGCAACCAAA GACGCACCTG
1451 CCGAAATCCG CGAACAATTC ATGCTGGCTG CGGAAAACAC GCTGAATATC
1501 TTTGCGCAAA AAGGCTATTT GGGATTGGAC GAATTTATTA CGTCAATAT
45 1551 CCCGAAAGGG CAGCAGGATA AGATGCAGGG CTATTCTAC GAAATGCTTT
1601 ACGGCGTGAT GAACGCTGCT TTGGATGAAA CCATACGCCG GTACGGCTTG
1651 CCCGAATGGC AGCAGGATGA AGCGCGGAAC CGTTTCCTGC TGCACAGTAT
1701 GGATGCCTAT ACGGGGCTGA CGGAATATCC CGCGCTATG CTGCTCCAGC
1751 TTGACGGGTT TTCCGAGGTG CGTTCCTCAG GTTTGCAGAT GACCCGTTCTG
1801 CCGGGTGCGC TTTTGGTCTA TCtgggctcg gtattgttgg TTTTGGgtac
50 1851 ggtatTtatg tTTTATGTGC GCGAAAAACG GGCGTGGgta tGTTTTCag
1901 aCGGCAAAAT CCGTTTTGCT ATGtCTTcgg CCcgagcgga ACGGGATTG
1951 cAGAaggaaT TTCCAAAACA CgtcgAGAGC CTGCAACgcg tcggcaaggA
2001 CttgaatCAT GACTga

```

This corresponds to the amino acid sequence <SEQ ID 336; ORF88ng-1>:

```

55 1  MSKSRIPTL LSRPWAFFS SMRFAVALLS LLGIASVIGT VLQONQPQTD
51  YLVKFGPFWT RIFDFLGLYD VYASAWFVVI MMFLVSTSL CLIRNVPPFW
101 REMKSFREKV KEKSLAAMRH SSLLDVKIAP EVAKRYLEVR GFQGKTVSRE
151 DGSVLIAAKK GTMNKWGYIF AQVALIVICL GGLIDSNLLL KLGMLAGRIV
201 PDNQAVYAKD FKPEILGAS NLSFRGNVNI SEGQSADVVF LNADNGMLVQ
60 251 DLPFEVKLKK FHIDFYNTGM PRDFASDIEV TDKATGEKLE RTIRVNHPLT
301 LHGITYIQAS FADGGSDLTF KAWNLRDASR EPVVLKATSI HQFPLEIGKH
351 KYRLEFDQFT SMNVEDMSEG AEREKSLKST LNDVRAVTQE GKKYTNIGPS
401 IVYRIRDAAG QAVEYKNYML PILQDKDYFW LTGTRSGLQQ QYRWLRIPLD
451 KQLKADTFMA LREFLKDEG RKRLVADATK DAPAEIREQF MLAAENTLNI
65 501 FAQKGYLGLD EFITSNIPKG QQDKMQGYFY EMLYGVMNAA LDETIRRYGL

```

551 PEWQQDEARN RFLHSMDAY TGLTEYPAPM LLQLDGFSEV RSSGLQMTRS
 601 PGALLVYLGS VLLVLGTVFM FYVREKRAWV LFS DGKIRFA MSSARSERDL
 651 QKEFPKHVES LQRLGKDLNH D*

ORF88ng-1 and ORF88-1 show 97.0% identity in 671 aa overlap:

```

5   orf88-1.pep  MSKSRRSPPLSRPWFAFFSSMRFAVALLSLGLIASVIGTVLQONQPOTDYLVKFGSFWA 60
      orf88ng-1  MSKSRI SPTLLSRPWFAFFSSMRFAVALLSLGLIASVIGTVLQONQPOTDYLVKFGPFWT 60

10  orf88-1.pep  QIFGFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120
      orf88ng-1  RIFDFLGLYDVYASAWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRH 120

15  orf88-1.pep  SLLDVKIAPEVAKRYLEVQGFQGTINREDGSLVIAAKKGTMNKGWYIFAQVALIVICL 180
      orf88ng-1  SLLDVKIAPEVAKRYLEVGRFQGTVSREDGSLVIAAKKGTMNKGWYIFAQVALIVICL 180

20  orf88-1.pep  GGLIDSNLLKLGMLTGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVV 240
      orf88ng-1  GGLIDSNLLKLGMLAGRIVPDNQAVYAKDFKPESILGASNLSFRGNVNISEGQSADVV 240

25  orf88-1.pep  LNADNGILVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT 300
      orf88ng-1  LNADNGMLVQDLPFEVKLKKFHIDFYNTGMPRDFASDIEVTDKATGEKLERTIRVNHPLT 300

30  orf88-1.pep  LHGITIYQASFADGGSDLTFKAWNLDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT 360
      orf88ng-1  LHGITIYQASFADGGSDLTFKAWNLRDASREPVLKATSIHQFPLEIGKHKYRLEFDQFT 360

35  orf88-1.pep  SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKTNIGPSIVYRIRDAAGQAVEYKNYML 420
      orf88ng-1  SMNVEDMSEGAEREKSLKSTLNDVRAVTQEGKKTNIGPSIVYRIRDAAGQAVEYKNYML 420

40  orf88-1.pep  PVLQEQDYFWITGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK 480
      orf88ng-1  PILQDKDYFWLTGTRSGLQQQYRWLRIPLDKQLKADTFMALREFLKDGEGRKRLVADATK 480

45  orf88-1.pep  GAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKEQQDKMQGYFYEMLYGVMNAA 540
      orf88ng-1  DAPAEIREQFMLAAENTLNIFAQKGYLGLDEFITSNIPKGQQDKMQGYFYEMLYGVMNAA 540

50  orf88-1.pep  LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS 600
      orf88ng-1  LDETIRRYGLPEWQQDEARNRFLHSMDAYTGLTEYPAPMLLQLDGFSEVRSSGLQMTRS 600

55  orf88-1.pep  PGALLVYLGSVLLVLGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES 660
      orf88ng-1  PGALLVYLGSVLLVLGTVLMFYVREKRAWVLFSDGKIRFAMSSARSERDLQKEFPKHVES 660

60  orf88-1.pep  LQRLGKDLNHD 671
      orf88ng-1  LQRLGKDLNHD 671
  
```

Furthermore, ORF88ng-1 shows homology with a hypothetical protein from *Aquifex aeolicus*:

```

55  gi|2984296 (AE000771) hypothetical protein [Aquifex aeolicus] Length = 537
      Score = 94.4 bits (231), Expect = 2e-18
      Identities = 91/334 (27%), Positives = 159/334 (47%), Gaps = 59/334 (17%)

60  Query: 16  FAFFSSMRFAVALLSLGLIASVIG-TVLQONQPOTDYLVKFGPFWTRIFDFLGLYDVYAS 74
      + F +S++ A+ ++ +LGI S++G T ++QNQ YL +FG L L DV+ S
      Sbjct: 80  YDFLASLKLAI FIMLVGLISMLGSTYIKQNSFEWYLDQFGYDVGIWIWKLWLNDFVHS 139

65  Query: 75  AWFVIMMFLVVSTSLCLIRNVPPFWREMKSFREKVKEKSLAAMRHSSLLDVKIAPEVAK 134
      ++++ ++ L V+ C I+ +P W++ S +E++ + A +H + VKI P+ K
      Sbjct: 140  WYYILFIVLLAVNLIFCSIKRLPRVWQAFS-KERILKLDHAEKHLKPITVKI-PDKDK 197

      Query: 135 --RYLEVGRFQGTVSREDGSLVIAAKKGTMNKGWYIFAQVALIVICLGGIDSNLLKL 192
      ++L +GF+ V E ++ A+KG ++ G ++AL+VI G LID
      Sbjct: 198  VLKFLKKGFK-VFVEEGNKLYVFAEKGRFSRLGVYITHIALLVIMAGALID----- 249
  
```

5 Query: 193 GMLAGRIVPDNQAVYAKDFKPESILGASNLSEFRGNVISEGQSAADVFLNADNGMLVQDL 252
 +I+G RG++ ++EG + DV+ + A+ L
 Sbjct: 250 -----AIVGV-----RGLIVAEGDTNDVMLVGAE--QKPYKL 280

10 Query: 253 PFEVKLKKFHIDFY---NTGMPRDF-----SDIEVTDKATGEKLER--TIRVNHPLT 300
 PF V L F I Y N + + FA SDIE+ + G K+E T++VN P
 Sbjct: 281 PFAVHLIDFRIKTYAEENPNVDKRFQAQVSSYESDIEIIN---GGKVEAKGTVKVNEPFD 337

Query: 301 LHGITYQASFA--DGGSDLTFKAWNLRDASREP 332
 ++QA++ DG S + + + A +P
 Sbjct: 338 FGRYRLFQATYGILDGTSGMGVIVVDRKKAHEDP 371

Based on this analysis, including the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 40

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 337>:

20 1 ATGATGAGTA ATAmAATGGM AAAAAAGGG TTTACATTGA TTGmGmTGAT
 51 GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT
 101 ATCmAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
 151 GYCGGTATCA ACAATATTTC CAAACAGTTT ATTTTGAAAA ATCCCCTGGA
 201 CGATAATCAG ACCATCGAGA ACAAAGTGA AATATTTGTC TCAGGCTATA
 251 AGATGAATCC GAAATTTGCC AAAAAaTATA GTGTTTCGGT AAAGTTTGTC
 25 GATAAGGAAA AATCAAGGGC ATACAGGTTG GTCGGCGTTC CGAAGGCGGG
 301 GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA
 351 AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCCTCA
 401 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAA
 451

This corresponds to the amino acid sequence <SEQ ID 338; ORF89>:

30 1 MMSNXMXQKG FTLLXXMIVV AILGIISVIA IPSYXSIEK GYQSOLYTEM
 51 XGINNISKQF ILKNPLDDNQ TIENKLEIFV SGYKMNPKIA KKYSVSVKVF
 101 DKEKSRAYRL VGVPKAGTGY TLSVWMNSVG DGYKCRDAAS AQAHLETLS
 151 DVGCEAFSNR KK*

Further work revealed the complete nucleotide sequence <SEQ ID 339>:

35 1 ATGATGAGTA ATAAATGGA AAAAAAGGG TTTACATTGA TTGAGATGAT
 51 GATAGTCGTC GCGATACTCG GCATTATCAG CGTCATTGCC ATACCTTCTT
 101 ATCAAAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
 151 GTCGGTATCA ACAATATTTC CAAACAGTTT ATTTTGAAAA ATCCCCTGGA
 201 CGATAATCAG ACCATCGAGA ACAAAGTGA AATATTTGTC TCAGGCTATA
 40 251 AGATGAATCC GAAATTTGCC AAAAAATATA GTGTTTCGGT AAAGTTTGTC
 301 GATAAGGAAA AATCAAGGGC ATACAGGTTG GTCGGCGTTC CGAAGGCGGG
 351 GACGGGTTAT ACTTTGTCGG TATGGATGAA CAGCGTGGGC GACGGATACA
 401 AATGCCGTGA TGCCGCTTCT GCCCAAGCCC ATTTGGAGAC CTTGTCCTCA
 451 GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAA

45 This corresponds to the amino acid sequence <SEQ ID 340; ORF89-1>:

1 MMSNKMEQKG FTLLIEMIVV AILGIISVIA IPSYQSYIEK GYQSOLYTEM
 51 VGINNISKQF ILKNPLDDNQ TIENKLEIFV SGYKMNPKIA KKYSVSVKVF
 101 DKEKSRAYRL VGVPKAGTGY TLSVWMNSVG DGYKCRDAAS AQAHLETLS
 151 DVGCEAFSNR KK*

50 Computer analysis of this amino acid sequence gave the following results:

Homology with Pile of *N. gonorrhoeae* (accession number Z69260).

ORF89 and Pile protein show 30% aa identity in 120a overlap:


```

orf89 8  QKGFTLIXXMIIVAILGIISVIAIPSYXSYIEKGYQSOLYTEMXGINNISKQFILKNPL- 66
        QKGFTLI  MIV+AI+GI++ +A+P+Y Y + S+ G + ++ L + +
File 5  QKGFTLIELMIVIAIVGILAAVALPAYQDYTARAQVSEAILLAEGOKSAVTEYYLNHGIW 64

5      orf89 67 -DDNQTIENTKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGYTLVSW 125
        DN + +G + KI KY SV + GV K G LS+W
File 65 PKDNTS-----AGVASSDKIKGKYVQSVTVAKGVVTAEMASTGVNKEIQGKKLSLW 115

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

10 ORF89 shows 83.3% identity over a 162aa overlap with an ORF (ORF89a) from strain A of *N. meningitidis*:

```

15      orf89.pep  MMSNXMXQKQFTLIXXMIIVAILGIISVIAIPSYXSYIEKGYQSOLYTEMXGINNISKQF
        orf89a    MMSNKMEQKQFTLIXXXXXAIXXXSVIXXXYSYIEKGYQSOLYTEMVGINNISKQX
                10      20      30      40      50      60

20      orf89.pep  ILKNPLDDNQTIENTKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGY
        orf89a    ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNVSVHFNEEKPRAYSLVGVPKTGTGY
                70      80      90      100     110     120

25      orf89.pep  TLSVWMNSVGDGYKCRDAASAQAHALETLSVDVGCEAFSNRKKX
        orf89a    TLSVWMNSVGDGYKCRDAASARAHLETLSVDVGCEAFSNRKKX
                130     140     150     160

```

The complete length ORF89a nucleotide sequence <SEQ ID 341> is:

```

30      1  ATGATGAGTA ATAAATGGA ACAAAAAGGG TTTACATTGA TTGNGANGNT
        51  NATNGNCNTC GCGATACNCN GCNTTANCAG CGTCATTNCN ATNNNTNCNT
        101  ATCNNAGTTA TATTGAAAAA GGCTATCAGT CCCAGCTTTA TACGGAGATG
        151  GTCGGTATCA ACAATATTTC CAAACAGTNT ATTTTGAAAA ATCCCTCGGA
        201  CGATAATCAG ACCATCAAGA GCAAACGGA AATATTGTGC TCAGGCATA
35      251  AGATGAATCC GAAAATTGCC GAAAAATATA ATGTTTCGGT GCATTTGTGC
        301  AATGAGGAAA AACCNAGGGC ATACAGCTTG GTCGGCGTTC CAAAGACGGG
        351  GACGGGTTAT ACTTTGTGCG TATGGATGAA CAGCGTGGGC GACGGATACA
        401  AATGCCGTGA TGCCGCTTCT GCCCGAGCCC ATTTGGAGAC CTTGTCCTCA
        451  GATGTCGGCT GTGAAGCCTT CTCTAATCGT AAAAAATAG

```

40 This encodes a protein having amino acid sequence <SEQ ID 342>:

```

        1  MMSNKMEQKG FTLIXXXXXX AIXXXSVIX XXXYXSYIEK GYQSOLYTEM
        51  VGINNISKQX ILKNPLDDNQ TIKSKLEIFV SGYKMNPKIA EKYNVSVHFN
        101  NEEKPRAYSL VGVPKTGTGY TLSVWMNSVG DGYKCRDAAS ARAHALETLS
        151  DVGCEAFSNR KK*

```

45 ORF89a and ORF89-1 show 83.3% identity in 162 aa overlap:

```

50      orf89a.pep  MMSNKMEQKQFTLIXXXXXAIXXXSVIXXXYSYIEKGYQSOLYTEMVGINNISKQX
        orf89-1    MMSNKMEQKQFTLIELMIVVAILGIISVIAIPSYQSYIEKGYQSOLYTEMVGINNISKQF
                10      20      30      40      50      60

55      orf89a.pep  ILKNPLDDNQTIKSKLEIFVSGYKMNPKIAEKYNVSVHFNEEKPRAYSLVGVPKTGTGY
        orf89-1    ILKNPLDDNQTIENTKLEIFVSGYKMNPKIAKKYSVSVKFVDKEKSRAYRLVGVPKAGTGY
                70      80      90      100     110     120

60      orf89a.pep  TLSVWMNSVGDGYKCRDAASARAHLETLSVDVGCEAFSNRKKX
        orf89-1    TLSVWMNSVGDGYKCRDAASAQAHALETLSVDVGCEAFSNRKKX
                130     140     150     160

```

ORF89 shows 84.6% identity over a 162aa overlap with a predicted ORF (ORF89.ng) from *N.*

[illegible]

	1	aTGATGAGCA	ATAAAATGGA	ACAAAAAGGG	TTTACATTGA	TTGAGATGAT
20	51	GATAGTTGTC	ACGATACTCG	GCATCATCAG	CGTCATTGCC	ATACCTTCTT
	101	ATCAGAGTTA	TATTGAAAAA	GGCTATCAGT	CCCAGCTTTA	TACGGAGATG
	151	GTCGGGTATCA	ACAATGTTCT	CAAAACAGTTT	ATTTTGAAAA	ATCCCCAGGA
	201	CGATAATGAT	ACCTTCAAGA	GCAAACTGAA	AATATTGTGC	TCAGGCTATA
	251	AGATGAATCC	GAAAAttgCC	AAAAAATATA	GTGTTTCGGT	aaggtttGTC
25	301	gatGCGGAAA	AACCAAGGGC	ATACAGGTTG	GTCGGCGTTC	CGAACGCGGG
	351	GACGGGTTAT	ACTTTGTCGG	TATGGATGAA	CAGCGTGGGC	GACGGATACA
	401	AATGCCGTGA	TGCCACTTCT	GCCCAGGCCT	ATTCCGACAC	CTTGTCGCGA
	451	GATAGCGGCT	GTGAAGCTTT	CTCTAATCGT	AAAAAATAG	

30 1 MMSNKMEQKG FTLIEMMIVV TILGIISVIA IPSYQSYIEK GYQSOLYTEM
51 VGINNVLKQF ILKNPQDDND TLKSKLKIFV SGYKMNPKIA KKYSVSVRFV
101 DAEKPRAYRL VGVPNAGTGY TLSVWMNSVG DGYKCRDATS AQAYSDTLSA
151 DSGCEAFSNR KK*

35 identity in 162 aa overlap:

		10	20	30	40	50	60
	orf89-1.pep	MMSNKMEQKGFTLIEMMIVVAILGIISVIAIPSYQSYSIEKGYQSQLYTEMVGINNISKQF					
		: : : : :					
40	orf89ng	MMSNKMEQKGFTLIEMMIVVTILGIISVIAIPSYQSYSIEKGYQSQLYTEMVGINNVLKQF					
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf89-1.pep	ILKNPLDDNQTIENKLEIFVSGYKMNPKIACKYSVSVKFVDKEKSRAYLRLGVGPAGTGY					
		: : : : :					
45	orf89ng	ILKNPQDDNDTLKSKLKFVSGYKMNPKIACKYSVSVRFDVDAEKPRAYRLGVGPNAGTGY					
		70	80	90	100	110	120
		130	140	150	160		
	orf89-1.pep	TLSVWMNSVGDGYKCRDAASAQAHALETSSDVGCEAFSRRKKX					
		: : : : :					
50	orf89ng	TLSVWMNSVGDGYKCRDATSAQAYSDTLSADSGCEAFSRRKKX					
		130	140	150	160		

BNSDOCID: <WO__9924578A2_1_>

ORF89-1 (13.6kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 11A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera gave a positive result in the ELISA test., confirming that

5 ORF89-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 41

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 345>:

```

1  ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
51  CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAGCCAA ATCCGTCAAA
101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
151 CGCCAAAAAG CCGAAGCCTA TGCATTCCC TATTTCGATT TCCAACGTAT
201 GACCGCATTG GCGGTCGGCA ACCCTGGsG CACCG.GTCC GACG.GCAAA
251 AACAAAGCGT GGCCn.AGAA TTTCAACCC...
```

This corresponds to the amino acid sequence <SEQ ID 346; ORF91>:

```

1  MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNTQVLS ILKNGDANTA
51  RQKAEAYAIP YDFQRM TAL AVGNPWXTXS DXQKQALAXE FQP...
```

Further work revealed the complete nucleotide sequence <SEQ ID 347>:

```

1  ATGAAAAAAT CCTCCCTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
51  CGGCATGGCA TTTGCCGCCC CTGCCGACGC GGTAAGCCAA ATCCGTCAAA
101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA ACGGCGATGC CAACACCGCT
151 CGCCAAAAAG CCGAAGCCTA TGCATTCCC TATTTCGATT TCCAACGTAT
201 GACCGCATTG GCGGTCGGCA ACCCTGGCG CACCGCGTCC GACGCGCAAA
251 AACAAAGCGT GGCCAAAGAA TTTCAACCCC TGCTGATCCG CACCTATTCC
301 GGCACGATGC TGAAATTAAA AAACGCCAAC GTCAACGTCA AAGACAATCC
351 CATCGTCAAT AAAGGCGGCA AAGAAATCAT CGTCCGCGCC GAAGTCGGCG
401 TACCCGGGCA AAAACCCGTC AACATGGACT TCACCACCTA CCAAAGCGGC
451 GGTAAATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
501 CGTGATCCGC AACCAATTCG GCGAAATTAT CAAAGCGAAA GCGGTGGACG
551 GACTGATTGC CGAGTTGAAA GCCAAAACG GCGGCAAATA A
```

30 This corresponds to the amino acid sequence <SEQ ID 348; ORF91-1>:

```

1  MKKSSLISAL GIGILSIGMA FAAPADAVSQ IRQNTQVLS ILKNGDANTA
51  RQKAEAYAIP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
101 GTMLKLKNNAN VNVKDNPIVN KKGKEIIVRA EVGVPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGASLVTYR NQFGEI IKAK GVDGLIAELK AKNGGK*
```

35 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF91 shows 92.4% identity over a 92aa overlap with an ORF (ORF91a) from strain A of *N.*

meningitidis:

```

40      orf91.pep      10      20      30      40      50      60
      MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNTQVLSILKNGDANTARQKAEAYAIP
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      orf91a          MKKSSFISALGIGILSIGMAFAAPADAVNQIRQNTQVLSILKSGDANTARQKAEAYAIP
                        10      20      30      40      50      60

45      orf91.pep      70      80      90
      YDFQRM TALAVGNPWXTXS DXQKQALAXE FQP
      |||||:|||||:|||||:|||||:|||||:|||||
      orf91a          YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYS GTMLKLKNNANVNVKDNPIVN
                        70      80      90      100      110      120
```

orf91a KGGKEIIVRAEVGVPQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI IKAK
130 140 150 160 170 180

The complete length ORF91a nucleotide sequence <SEQ ID 349> is:

```

5      1 ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
      51 CGGCATGGCA TTTGCCGCC CTGCCGACGC GGTAACCAA ATCCGTCAAA
     101 ACGCCACTCA AGTATTGAGC ATCTTAAAAA GCGGTGATGC CAACACCGCC
     151 CGCCAAAAAG CCGAAGCCTA TGCGATTCCC TATTTGATT TCCAACGTAT
     201 GACCGCATTG GCGGTCGGCA ACCCTTGGCG CACCGCGTCC GACGCGCAAA
10     251 AACAAAGCGT GGCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTCC
     301 GGCACGATGC TGAAATTAAA AAACGCCAAC GTCAACGTCA AAGACAATCC
     351 CATCGTCAAT AAAGCGGCA AAGAAATCAT CGTCCGCCCC GAAGTCGGCG
     401 TACCCGGGCA AAAACCGGTC AACATGGACT TCACCACCTA CCAAAGCGGC
     451 GGTAAATACC GTACCTACAA CGTCGCCATC GAAGGCGCGA GCCTGGTTAC
15     501 CGTGTACCGC AACCAATTCG GCGAAATTAT CAAAGCGAAA GGCGTGGACG
     551 GACTGATTGC CGAGTTGAAG GCTAAAAACG GCAGCAAGTA A

```

This encodes a protein having amino acid sequence <SEQ ID 350>:

```

      1 MKKSSFISAL GIGILSIGMA FAAPADAVNQ IRQNATQVLS ILKSGDANTA
     51 RQKAEAYAIP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
20    101 GTMLKLKNNAN VNVKDNPIVN KGGKEIIVRA EVGVPQKPV NMDFTTYQSG
     151 GKYRTYNVAI EGASLVTYR NQFGEI IKAK GVDGLIAELK AKNGSK*

```

ORF91a and ORF91-1 show 98.0% identity in 196 aa overlap:

```

25      10      20      30      40      50      60
    orf91a.pep MKKSSFISALGIGILSIGMAFAAPADAVNQIRQNATQVLSILKSGDANTARQKAEAYAIP
              ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
    orf91-1    MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP
              10      20      30      40      50      60

      70      80      90      100     110     120
    orf91a.pep YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGMTMLKLKNNANVNVKDNPIVN
              |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
    orf91-1    YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGMTMLKLKNNANVNVKDNPIVN
              70      80      90      100     110     120

35      130     140     150     160     170     180
    orf91a.pep KGGKEIIVRAEVGVPQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI IKAK
              |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
    orf91-1    KGGKEIIVRAEVGVPQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI IKAK
              130     140     150     160     170     180

40      190
    orf91a.pep GVDGLIAELKAKNGSKX
              |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
    orf91-1    GVDGLIAELKAKNGGKX
              190
45

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF91 shows 84.8% identity over a 92aa overlap with a predicted ORF (ORF91.ng) from *N. gonorrhoeae*:

```

50      orf91.pep MKKSSLISALGIGILSIGMAFAAPADAVSQIRQNATQVLSILKNGDANTARQKAEAYAIP 60
           :||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      orf91ng    VKKSSFISALGIGILSIGMAFASPADAVGQIRQNATQVLTILKSGDAASARPKAEAYAVP 60

           orf91.pep YDFQRM TALAVGNPWXTXSDXQKQALAXEFQP 93
           |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
55      orf91ng    YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYSGMTMLKFNATVNVKDNPIVN 120

```

The complete length ORF91ng nucleotide sequence <SEQ ID 351> is predicted to encode a protein having amino acid sequence <SEQ ID 352>:

-230-

```

1 VKKSSFISAL GIGILSIGMA FASPADAVGQ IRONATQVLT ILKSGDAASA
51 RPKAEAYAVP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGTSLVTVYR NQFGEI IKAK GIDGLIAELK AKNGGK*

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 353>:

```

1 ATGAAAAAAT CCTCCTTCAT CAGCGCATTG GGCATCGGTA TTTTGAGCAT
51 CGGCATGGCA TTTGCCTCCC CGGCCGACGC AGTGGGACAA ATCCGCCAAA
101 ACGCCACACA GGTTTGTACC ATCCTCAAAA GCGGCGACGC GGCTTCTGCA
151 CGCCCAAAAG CCGAAGCCTA TGCGGTTCCC TATTTCGATT TCCAACGTAT
201 GACCGCATTG GCGGTCGGCA ACCCTTGGCG TACCGCGTCC GACGCGCAAA
251 AACAAGCGTT GCCCAAAGAA TTTCAAACCC TGCTGATCCG CACCTATTC
301 GGCACGATGC TGAAATTCAA AAACGCGACC GTCAACGTCA AAGACAATCC
351 CATCGTCAAT AAGGGCGGCA AGGAAATCGT CGTCCGTGCC GAAGTCGGCA
401 TCCCGGTCA GAAGCCGTC AATATGGA CTACCACTA CCAAAGCGGC
15 451 GGCAATACC GTACCTACAA CGTCGCCATC GAAGGCACGA GCCTGGTTAC
501 CGGTACCGC AACCAATTCG GCGAAATCAT CAAAGCCAAA GGCATCGACG
551 GGCTGATTGC CGAGTTGAAA GCCAAAACG GCGGCAATA A

```

This corresponds to the amino acid sequence <SEQ ID 354; ORF91ng-1>:

```

1 MKKSSFISAL GIGILSIGMA FASPADAVGQ IRONATQVLT ILKSGDAASA
20 51 RPKAEAYAVP YDFQRM TAL AVGNPWRTAS DAQKQALAKE FQTLIRTYS
101 GTMLKFKNAT VNVKDNPIVN KGGKEIVVRA EVGIPGQKPV NMDFTTYQSG
151 GKYRTYNVAI EGTSLVTVYR NQFGEI IKAK GIDGLIAELK AKNGGK*

```

ORF91ng-1 and ORF91-1 show 92.3% identity in 196 aa overlap:

```

25 orf91-1.pep 10 20 30 40 50 60
    MKKSSLISALGIGILSIGMAFAPADAVSQIRONATQVLSILKNGDANTARQKAEAYAIP
    orf91ng-1  MKKSSFISALGIGILSIGMAFASPADAVGQIRONATQVLTILKSGDAASARPKAEAYAVP
                10 20 30 40 50 60
30 orf91-1.pep 70 80 90 100 110 120
    YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYS GTMLKLNANVNVKDNPIVN
    orf91ng-1  YDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYS GTMLKFKNATVNVKDNPIVN
                70 80 90 100 110 120
35 orf91-1.pep 130 140 150 160 170 180
    KGGKEIIVRAEVGVPGQKPVNMDFTTYQSGGKYRTYNVAIEGASLVTVYRNQFGEI IKAK
    orf91ng-1  KGGKEIVVRAEVGIPGQKPVNMDFTTYQSGGKYRTYNVAIEGTSLVTVYRNQFGEI IKAK
                130 140 150 160 170 180
40 orf91-1.pep 190
    GVDGLIAELKAKNGGKX
    orf91ng-1  GIDGLIAELKAKNGGKX
                190
45

```

In addition, ORF91ng-1 shows homology to a hypothetical *E. coli* protein:

```

50 sp|P45390|YRBC_ECOLI HYPOTHETICAL 24.0 KD PROTEIN IN MURA-RPON INTERGENIC
    REGION PRECURSOR (F211) >gi|606130 (U18997) ORF_f211 [Escherichia coli]
    >gi|1789583 (AE000399) hypothetical 24.0 kD protein in mur2-rpoN intergenic
    region [Escherichia coli] length = 211
    Score = 70.6 bits (170), Expect = 6e-12
    Identities = 42/137 (30%), Positives = 76/137 (54%), Gaps = 6/137 (4%)
55 Query: 59 VPYDFQRM TALAVGNPWRTASDAQKQALAKEFQTLIRTYS GTMLKFKNATVNVKDNPI 118
    +PY + AL +G +++A+ AQ++A F+ L + Y + + T + P
    Sbjct: 65 LPYVQVKYAGALVLGQYYKSATPAQREAYFAAFREYLKQAYGQALAMYHGQTYQIA--PE 122
60 Query: 119 VNKGGKEIV-VRAEVGIP-GQKPVNMDFTTYQSG--GKYRTYNVAIEGTSLVTVYRNQFG 174
    G K IV +R + P G+ PV +DF ++ G ++ Y++ EG S++T +N++G
    Sbjct: 123 QPLGDKTIVPIRVTIIDPNGRPPVRLDFQWRKNSQTGNWQAYDMIAEGVSMITTKQNEWG 182

```

Query: 175 EIIKAKGIDGLIAELKA 191
 +++ KGIDGL A+LK+
 Sbjct: 183 TLLRTKGIDGLTAQLKS 199

Based on this analysis, including the presence of a putative leader sequence in the gonococcal
 5 protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes,
 could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 42

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 355>:

```

10      1  ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTCAAC
      51  CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACTCAAAAC GAAACCGCTA
     101  TGATCACGCA TACCCTCATC TCAAAATACA GTTTTGnnn nnnnnnnnnn
     151  nnnnnnnnnn nnGCCATAAA AAGCAAAGGG ATGGACATT TTGCCGTCAT
     201  CGACCATCAG GAAGCCGCAC GCCGAAACGG CTTAACGATG CAGCCGGCAA
     251  AAGTCATCGT CTTGGGCACG CCCAAAGCCG GCACGCCGCT GATGGTCAAA
     301  GACCCCGCCT TCGCCCTGCA ACTGCCCTA CGCGTCCTCG TTACCGAAAC
     351  GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
     401  GCAGCCGCAT CGGTTTCGAC GAAGTGCAA ACACTTTGGC AAACGCCGAA
     451  AAACGTATAC AAAAAACCGT AGGCGAATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 356; ORF97>:

```

20      1  MKHILPLIAA SALCISTASA HPASEPSTQN ETAMITHTLI SKYSFGXXXX
      51  XXXXAISKSG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
     101  DPAFALQLPL RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
     151  KLIQKTVE*
  
```

Further work revealed the complete nucleotide sequence <SEQ ID 357>:

```

25      1  ATGAAACACA TACTCCCCCT GATTGCCGCA TCCGCACTCT GCATTTCAAC
      51  CGCTTCGGCA CATCCTGCCA GCGAACCGTC CACCCAAAAC GAAACCGCTA
     101  TGACCACGCA TACCCTCACC TCAAAATACA GTTTTGACGA AACCCTCAGC
     151  CGCCTTGAAA CCGCCATAAA AAGCAAAGGG ATGGACATT TTGCCGTCAT
     201  CGACCATCAG GAAGCCGCCC GCCGAAACGG CTTAACGATG CAGCCGGCAA
     251  AAGTCATCGT CTTGGGCACG CCCAAAGCCG GCACGCCGCT GATGGTCAAA
     301  GACCCCGCCT TCGCCCTGCA ACTGCCCTA CGCGTCCTCG TTACCGAAAC
     351  GGACGGCAAA GTACGCGCCG CCTATACCGA TACGCGCGCC CTCATCGCCG
     401  GCAGCCGCAT CGGTTTCGAC GAAGTGCAA ACACTTTGGC AAACGCCGAA
     451  AAACGTATAC AAAAAACCGT AGGCGAATAA
  
```

35 This corresponds to the amino acid sequence <SEQ ID 358; ORF97-1>:

```

      1  MKHILPLIAA SALCISTASA HPASEPSTQN ETAMTHTLT SKYSFDETVS
      51  RLETAISKSG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
     101  DPAFALQLPL RVLVTETDGK VRAAYTDTRA LIAGSRIGFD EVANTLANAE
     151  KLIQKTVE*
  
```

40 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF97 shows 88.7% identity over a 159aa overlap with an ORF (ORF97a) from strain A of *N.meningitidis*:

```

45      orf97.pep      10      20      30      40      50      60
      MKHILPLIAASALCISTASAHHPASEPSTQNETAMITHTLISKYSFGXXXXXXXXXAISKSG
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      orf97a      MXHILPLXXASALCISTASXHPASEPQTQNETAMTHTLTISKYSFDETVSRLATAISKSG
                  10      20      30      40      50      60
  
```

-232-

		70	80	90	100	110	120
5	orf97.pep	MDIFAVIDHQEARRNGLTMQPAKVIVFGT	PKAGTPLMVKDPAFALQLPLRVLTETD	GK			
	orf97a	MDIFAVIDHQEARRNGLTMQPAKVIVFGT	PKAGTPLMVKDPAFALQLPLRVXVTETD	GK			
		70	80	90	100	110	120
10	orf97.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKT	VGEX				
	orf97a	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKT	IGEX				
		130	140	150	160		

The complete length ORF97a nucleotide sequence <SEQ ID 359> is:

	1	ATGANACACA	TACTCCCCCT	GANTGNCGCA	TCCGCACTCT	GCATTTCAAC
15	51	CGCTTCGGNN	CATCCTGCCA	GCGAACCGCA	AACCCAAAAC	GAAACCGCTA
	101	TGACCACGCA	TACCTCACC	TCAAATACA	GTTTGTACGA	AACCGTCAGC
	151	CGCCTTGAAA	CCGCCATAAA	AAGCAAAGGG	ATGGACATTT	TTGCCGTCAT
	201	CGACCATCAG	GAAGCCGCC	GCCGAAACGG	CTTAACGATG	CAGCCGGCAA
	251	AAGTCATCGT	CTTCGGCAGC	CCCAAAGCCG	GTACGCCGCT	GATGGTCAAA
20	301	GACCCCGCCT	TCGCCCTGCA	ACTGCCCTG	CGCGTCNTCG	TTACCGAAAC
	351	GGACGGCAAA	GTACGCGCCG	CCTATACCGA	TACGCGCGCC	CTCATCGCCG
	401	GCAGCCGCAT	CGGTTTCGAC	GAAGTGGCAA	ACACTTTGGC	AAACGCCGAA
	451	AAACTGATAC	AAAAAACCAT	AGGCGAATAA		

This encodes a protein having amino acid sequence <SEQ ID 360>:

	1	<u>MXHILPLXXA</u>	<u>SALCISTASX</u>	HPASEPQTQN	ETAMTHTLT	SKYSFDETVS
25	51	RLETAIKSKG	MDIFAVIDHQ	EAARRNGLTM	QPAKVIVFGT	PKAGTPLMVK
	101	DPAFALQLPL	RVXVTETDGK	VRAAYTDTRA	LIAGSRIGFD	EVANTLANAE
	151	KLIQKTIGE*				

ORF97a and ORF97-1 show 95.6% identity in 159 aa overlap:

		10	20	30	40	50	60
30	orf97a.pep	MXHILPLXXASALCISTASXHPASEPQTQN	ETAMTHTLT	SKYSFDETVSRLETAIKSKG			
	orf97-1	MXHILPLIAASALCISTASAHASEPSTQNETAMTHTLT	SKYSFDETVSRLETAIKSKG				
		10	20	30	40	50	60
35	orf97a.pep	MDIFAVIDHQEARRNGLTMQPAKVIVFGT	PKAGTPLMVKDPAFALQLPLRVXVTETD	GK			
	orf97-1	MDIFAVIDHQEARRNGLTMQPAKVIVFGT	PKAGTPLMVKDPAFALQLPLRVLTETD	GK			
		70	80	90	100	110	120
40	orf97a.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKT	IGEX				
	orf97-1	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKT	VGEX				
45		130	140	150	160		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF97 shows 88.1% identity over a 159aa overlap with a predicted ORF (ORF97.ng) from *N.gonorrhoeae*:

50	orf97.pep	MXHILPLIAASALCISTASAHASEPSTQNETAMITHTLISKYSFGXXXXXXXA	IKSKG	60
	orf97ng	MXHILPPIAASAFICISTASAHAGKPTQNETAMTHTLTISKYSFDETVSRLETAIKSKG		60
55	orf97.pep	MDIFAVIDHQEARRNGLTMQPAKVIVFGT	PKAGTPLMVKDPAFALQLPLRVLTETD	GK 120
	orf97ng	MDIFAVIDHQEARRNGLTMQPAKVIVFGT	PKAGTPLMVKDPAFALQLPLRVLTETD	GK 120
	orf97.pep	VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKT	VGEX	159
60	orf97ng	VRTAYTDTRALIVGSRISFDEVANTLANAEKLIQKT	VGEX	159

The complete length ORF97ng nucleotide sequence <SEQ ID 361> is predicted to encode a protein having amino acid sequence <SEQ ID 362>:

```

      1 MKHILPPIAA SAFCISTASA HPAGKPPTON ETAMTHTLT SKYSFDETVS
      51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
5    101 DPAFALQLPL RVLVTETDGK VRTAYTDTRA LIVGSRISFD EVANTLANAE
      151 KLIQKTVGE*

```

Further work revealed the complete nucleotide sequence <SEQ ID 363>:

```

      1 ATGAAACACA TACTCCCcct gatcgccgca TccgcactCT GCATTTCAC
      51 CGCTTCGGCA CACCTGCGG GCAAACCGCC CACCCAAAAC GAAACCGCTA
     101 TGACCACGCA CACCTCACC TCGAAATACA GTTTTGACGA AACCGTCAGC
     151 CGCCTTGAAA CCGCCATAAA AAGCAAAGGG ATGGACATT TTGCCGTCAT
     201 CGACCATCAG GAAGCGGCAC GCCGAAACGG CCTGACCATG CAGCCGGCAA
     251 AAGTCATCGT CTTCCGCACG CCCAAGGCCG GTACGCCgct GATGGTCAAA
     301 GACCCCGCCT TCGCCCTGCA ACTGCCCTG CGCGTCCTCG TTACCGAAAC
     15 351 GGACGGCAA GTACGCACCG CCTATACCGA TACGCGCGCC CTCATCGTCG
     401 GCAGCCGCAT CAGTTTCGAC GAAGTGGCAA ACACTTTGGC AAACGCCGAA
     451 AAATGATAC AAAAAACCGT AGGCGAATAA

```

This corresponds to the amino acid sequence <SEQ ID 364; ORF97ng-1>:

```

      1 MKHILPLIAA SALCISTASA HPAGKPPTON ETAMTHTLT SKYSFDETVS
     20 51 RLETAIKSKG MDIFAVIDHQ EAARRNGLTM QPAKVIVFGT PKAGTPLMVK
     101 DPAFALQLPL RVLVTETDGK VRTAYTDTRA LIVGSRISFD EVANTLANAE
     151 KLIQKTVGE*

```

ORF97ng-1 and ORF97-1 show 96.2% identity in 159 aa overlap:

```

25      10      20      30      40      50      60
      orf97-1.pep MKHILPLIAASALCISTASAHASEPSTONETAMTHTLT SKYSFDETVSRLETAIKSKG
      orf97ng-1   MKHILPLIAASALCISTASAHAGKPPTONETAMTHTLT SKYSFDETVSRLETAIKSKG
                10      20      30      40      50      60
30      70      80      90      100     110     120
      orf97-1.pep MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK
      orf97ng-1   MDIFAVIDHQEAARRNGLTMQPAKVIVFGTPKAGTPLMVKDPAFALQLPLRVLVTETDGK
                70      80      90      100     110     120
35      130     140     150     160
      orf97-1.pep VRAAYTDTRALIAGSRIGFDEVANTLANAEKLIQKTVGEX
      orf97ng-1   VRTAYTDTRALIVGSRISFDEVANTLANAEKLIQKTVGEX
                130     140     150     160
40

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF97-1 (15.3kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 12A & 12B show, respectively, the results of affinity purification of the GST-fusion and His-fusion proteins. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western Blot (Figure 12C), ELISA (positive result), and FACS analysis (Figure 12D). These experiments confirm that ORF97-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 12E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF97-1.

Example 43

The following DNA, believed to be complete, sequence was identified in *N.meningitidis* <SEQ ID 365>:

```

5      1 ATGGCTTTTA TTACGCGCTT ATTCAAAGC AGTAAATGGC TGATTGTGCC
      51 GCTGATGCTC CCCGCCTTTC AGAATGTGGC GCGGAGGGG ATAGATGTGA
     101 GCCGTGCCGA AGCGAGGATA ACCGACGGCG GGCAGCTTTC CATCAGCAGC
     151 CGCTTCCAAA CCGAGCTGCC CGACCAGCTC CAACAGGCGT TCGCGCGGGG
     201 CGTGCCGCTC AACTTTACCT TAAGCTGGCA GCTTTCGCC CCGATAATCG
    10      251 CTTCTTATCG GTTTAAATTG GGGCAACTGA TTGGCGATGA CGACAATATT
      301 GACTACAAAC TGAGTTTCCA TCCGCTGACC AaACGCTACC GCGTTACCgT
      351 CGGCGCGGTT TCGACAGACT ACGACACCTT GGATGCGGCA TTGCGCGCGA
     401 CCGGCGCGGT TGCCAACCTG AAAGTCCTGA ACAAAGGCGC GCTGTCCGGT
     451 GCGGAAGCAG GGGAAACCAA GCGGAAATC CGCCTGACGC TGTCCACTTC
    15      501 AAAACTGCCC AAGCCTTTTC AAATCAATGC ATTGACTTCT CAAAACTGGC
      551 ATTTGGATTC GGGTTGGAAA CCTCTAAACA TCATCGGGAA CAAATAA
  
```

This corresponds to the amino acid sequence <SEQ ID 366; ORF106>:

```

      1 MAFITRLFKS SKWLIVPLML PAFQNVAAEG IDVSRAEARI TDGGQLSISS
     51 RFQTELPDQL QQALRRGVPL NFTLSWQLSA PIIASYRFLK GQLIGDDDDNI
    20      101 DYKLSFHPLT KRYRVTVGAF STDYDTLDAA LRATGAVANW KVLNKGALSG
      151 AEAGETKAEI RLTLSTSKLP KPFQINALTS QNWHLD SGWK PLNIIGNK*
  
```

Further work revealed the following DNA sequence <SEQ ID 367>:

```

      1 ATGGCTTTTA TTACGCGCTT ATTCAAAGC AGTAAATGGC TGATTGTGCC
     51 GCTGATGCTC CCCGCCTTTC AGAATGTGGC GCGGAGGGG ATAGATGTGA
    25      101 GCCGTGCCGA AGCGAGGATA ACCGACGGCG GGCAGCTTTC CATCAGCAGC
     151 CGCTTCCAAA CCGAGCTGCC CGACCAGCTC CAACAGGCGT TCGCGCGGGG
     201 CGTGCCGCTC AACTTTACCT TAAGCTGGCA GCTTTCGCC CCGATAATCG
     251 CTTCTTATCG GTTTAAATTG GGGCAACTGA TTGGCGATGA CGACAATATT
     301 GACTACAAAC TGAGTTTCCA TCCGCTGACC AACCCTGACC GCGTTACCGT
    30      351 CCGCGCGGTT TCGACAGACT ACGACACCTT GGATGCGGCA TTGCGCGCGA
     401 CCGGCGCGGT TGCCAACCTG AAAGTCCTGA ACAAAGGCGC GCTGTCCGGT
     451 GCGGAAGCAG GGGAAACCAA GCGGAAATC CGCCTGACGC TGTCCACTTC
     501 AAAACTGCCC AAGCCTTTTC AAATCAATGC ATTGACTTCT CAAAACTGGC
    35      551 ATTTGGATTC GGGTTGGAAA CCTCTAAACA TCATCGGGAA CAAATAA
  
```

35 This corresponds to the amino acid sequence <SEQ ID 368; ORF106-1>:

```

      1 MAFITRLFKS SKWLIVPLML PAFQNVAAEG IDVSRAEARI TDGGQLSISS
     51 RFQTELPDQL QQALRRGVPL NFTLSWQLSA PIIASYRFLK GQLIGDDDDNI
    40      101 DYKLSFHPLT NRYRVTVGAF STDYDTLDAA LRATGAVANW KVLNKGALSG
      151 AEAGETKAEI RLTLSTSKLP KPFQINALTS QNWHLD SGWK PLNIIGNK*
  
```

40 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF106 shows 87.4% identity over a 199aa overlap with an ORF (ORF106a) from strain A of *N.meningitidis*:

```

45      orf106.pep      10      20      30      40      50      59
      MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAEARITDGGQLSISSRFQTELPDQ
      ||||| |||:: ||::: ||||| ||||| ||||| ||||| ||||| |||||
      orf106a          MAFITRLFKSIKWLVLPLMLSVLPDAAAEGIDVSRAEARIXDGGQLSXSRFQTELPDQ
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      10      20      30      40      50      60

50      orf106.pep      60      70      80      90      100      110      119
      LQQALRRGVPLNFTLSWQLSAPIIASYRFLKGLIGDDDDNIDYKLSFHPLTKRYRVTGVA
      ||| ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  
```

The complete length ORF106a nucleotide sequence <SEQ ID 369> is:

	1	ATGGCTTTTA	TTACGCGCTT	ATTCAAAAGC	ATTAACAAT	GGCTTGTGCT
20	51	GCTGCCGATG	CTTCCGTTT	TGCCGGACGC	GGCGCGGAG	GGGATAGATG
	101	TGAGCCGCGC	CGAAGCGAGG	ATAANCGACG	CGCGGCAGCT	TCGCATNAGN
	151	AGCCGCTTCC	AAACCGAGCT	GCCCGACCA	GCCGNNNGNG	CCGCGCCG
	201	GGCGGTGNCG	CTCAACTNTA	CCTTAAAGNTG	GCAGCTTTCC	GCCCGGATAA
	251	TCGCTTCTTA	TCGGTTTNA	TTGGGGCAAC	TGATTGGCGA	TGACGACNAT
25	301	ATTGACTACA	AACTGAGTTT	CCATCCGCTG	ACCAACCGCT	ACCGCGTTAC
	351	CGTCGGCGCG	TTTTCGACAG	ANTACGACAC	CTTGATGCG	GATTGCGCG
	401	CGACCGGCGC	GGTTGCCAAC	TGGAAGTCC	TGAACAAAGG	CGCGCTGTCC
	451	GGTGCGGAAG	CAGGGGAAAC	CAAGGCGGAA	ATCCGCGCTGA	CGCTGTCCAC
	501	TTCAAACCTG	CCCCAGCCTT	TTCAAATCAA	TGCATTGACT	TCTCAAACCT
	551	GGCATTGGA	TCGGGTTGG	AAACCTCTAA	ACATCATCGG	GAACAAATAA

```

1  MAFITRLFKS IKQWLVLPLM LSVLPDAAAE GIDVSRAEAR IXDGGQLSXX
51 SRFQTELPDQ LQXAXXRGVX LNXTLXWQLS APIIASYRFX LGQLIGDDDX
101 IDYKLSFHPL TNRYRVTVGA FSTXYDTLDA ALRATGAVAN WKVLNKGALS
151 GAEAGETKAE IRLTLSTSKL PKPFOINALT SQNWHLDSGW KPLNIIGNK*

```

ORF106 shows 90.5% identity over a 199aa overlap with a predicted ORF (ORF106.ng) from *N. gonorrhoeae*:

40	orf106.pep	MAFITRLFKSSK-WLIVPLMLPAFQNVAAEGIDVSRAEARITDGGQLSISRRFQTELPDQ	59
	orf106ng	MAFITRLFKSIKQWLVLLPILSVLPDAAAEGIAATRAEARITDGGRLSISRRFQTELPDQ	60
45	orf106.pep	LQQALRRGVPLNFTLSWQLSAPIIASYRFLKGQLIGDDDNIDYKLSFHPLTKRYRVTVGA	119
	orf106ng	LQQALRRGVPLNFTLSWQLSAPTIASYRFLKGQLIGDDDNIDYKLSFHPLTNRYRVTVGA	120
50	orf106.pep	FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLPKPFQINALT	179
	orf106ng	FSTDYDTLDAALRATGAVANWKVLNKGALSGAEAGETKAEIRLTLSTSKLPKPFQINALT	180
	orf106.pep	SONWHLDSGWKPLNIIGNK	198
	orf106ng	SONWHLDSGWKPLNIIGNK	199

Due to the K→N substitution at residue 111, the homology between ORF106ng and ORF106-1 is 91.0% over the same 199 aa overlap.

The complete length ORF106ng nucleotide sequence <SEQ ID 371> is:

```

1  ATGGCTTTTA TTACGCGCTT ATTCAAAAGC ATTAACAAT GGCTTGTGCT
51  GTTGCCGATA CTCTCCGTTT TGCCGGACGC GCGGCGGAG GGCATTGCCG
101 CGACCCGCGC CGAAGCGAGG ATAACCGACG GCGGGCGGCT TTCCATCAGC
151 AGCCGCTTCC AAACCGAGCT GCCCGACCAG CTCCAACAGG CGTTGCGCCG
201 GGGCGTACCG CTCAACTTTA CCTTAAGCTG GCAGCTTTCG GCCCCGACAA
251 TCGCTTCTTA TCGGTTTAAA TTGGGGCAAC TGATTGGCGA TGACGACAA
301 ATTGACTACA AACTAAGTTT CCATCCGCTG ACCAACCGCT ACCGCGTTAC
351 CGTCGGCGCA TTTTCCACCG ATTACGACAC TTTGGATGCG GCATTGCGCG
401 CGACCCGCGC GGTGCCCAAC TGGAAAGTCC TGAACAAAGG CGCGTTGTCC
451 GGTGCGGAAG CAGGGGAAAC CAAGGCGGAA ATCCGCTGA CGCTGTCCAC
501 TTCAAAACTG CCCAAGCCTT TCCAAATCAA CGCATTGACT TCTCAAACT
551 GGCATTGGA TTCGGGTTGG AAACCTCTAA ACATCATCGG GAACAAATAA

```

This encodes a protein having amino acid sequence <SEQ ID 372>:

```

15 1  MAFITRLFKS IKQWLVLPI LSVLPDAAE GIAATRAEAR ITDGGRLSIS
51  SRFQTELPDQ LQQALRRGVP LNFTLSWQLS APTIASYRFK LGQLIGDDDN
101 IDYKLSFHPL TNRYRVTGVA FSTDYDTLDA ALRATGAVAN WKVLNKGALS
151 GAEAGETKAE IRLTSTSKL PKPFQINALT SQNWHLDSGW KPLNIIGNK*

```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF106-1 (18kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 13A shows the results of affinity purification of the His-fusion protein, and Figure 13B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 13C) These experiments confirm that ORF106-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 44

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 373>:

```

30 1  ATGGACACAA AAGAAATCCT CGG.TACGCG GcAGGcTCGA TCGGCAGCGC
51  GGTTTTAGCC GTCATCATCc TGCCGCTGCT GTCGTGGTAT TTCCCCGCGC
101 ACGACATCGG GCGCATCGTG CTGATGCAGA CCGCGCGGGG GCTgACGGTG
151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
35 201 CACCGCCGAC AAAGACAcCT TGTTCAAAC CCTGTTCCTG CCGCCGCTGC
251 TGTCTGCCGC GCGATAGACC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
301 TCTGAAATCC TGTTTTCAC TCGAGATGCC gCCGCGGCa TCGGGCTGGT
351 GCTGTTTGAA CtGAGCTTCC TGCCCATCCG cTTTCTCTTA CTGGTTTTGC
401 GTATGGAAGG ACGCGCCcTT GCCTTTTCGT CCGCGCAACT CGTGCCcAAG
40 451 CTCGCCATCC TGCTGTGT.T GCCGCTGACG GTCGGGCTGC TGCACTTTCC
501 AGCGAACACC GCCGTCTGA CCGCGTTTA CCGCTGGCA AACCTTGCCG
551 CCGCCGCCTT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCGG
601 CACGCACCGT TTTGCGCCGC GTCCTGCAC CGGGGG.TGC GCTACGGCAT
651 ACCGATCGCA CTGAGCAGCA TCGCCTATTG GGGGCTGGCA TCCGCGACC
45 701 GTTTGTTTCT GAAAAAATAT GCCGGCCTGG AACAGCTCGG CGTTTATTTC
751 ATGGGTATTT CGTTCGGCGG GCGGCATTA TTGTTCCAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTTCCGCGC AATCGAAGAA AACGCCCCGC
851 CCGCTCGCCT CTCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCCCTC
901 GCCCTCTGC.TGACCGGCAT TTCTCGCCC CTTGCCTCCC TCCTGCTGCC
50 951 GGAAACTAC GCCGCCGTCC GGTTCATCGT CGTATCGTGT ATG.TGCCGC

```

-237-

5
1001 CGCTGTTTTC CACGCTGGCG GAAATCAGCG GCATCGGTTT GAACGTCGTT
1051 CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CCTGCTGCTG CTGGGGCTTG ACCGTGCCGT ACCGGCGAGG CCGCC.GGCG
1151 CGGCGGTTGC CTGTGCCGCC TCATTCTGGC TGTTTTTTGC CTTCAAGACC
1201 GAAAGCTCyt GCCGCCTGTG GCAGCCGCTC AAACGCCTGC CGCTTTATCT
1251 GCACACATTG TTCTGCCTGA CCTCCTCGGC GGCCTACACC TGCTTCGGCA
1301 CGCCGGCAAA CTATCCCCTG TTGCGCGCG TATGGGCGGC ATATCTGGCA
1351 GGCTGCATCC TGCGCCACCG GAAAGATTG CACAACTGT TTCATTATTT
1401 GAAAAACAA GGTTCCTCAT TATGA

10 This corresponds to the amino acid sequence <SEQ ID 374; ORF10>:

15
20
25
30
35
40
45
1 MDTKEILXYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YVREYYATAD KDTLFKTLFL PPLLSAAAIA ALLLSRPSLP
101 SEILFSLDDA AAGIGLVLE LSLPIRELL LVLRMEGRAL AFSSAQLVPK
151 LAIIILLXPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR
201 HAPFSPAVLH RGXYRGIPIA LSSIAYWGLA SADRLFLKKY AGLEQLGVYS
251 MGISFGGAAL LFQSIYSTVW TPYIFRAIEE NAPPARLSAT AESAAALLAS
301 ALCLTGIFSP LASLLLPENY AAVRFIVVSC MLPLFCTLA EISGIGLNVV
351 RKTRPIALAT LGALANLLL LGLDRAVPAR PXGAAVACAA SFWLFFAFKT
401 ESSCRLWQPL KRLPLYLHTL FCLTSSAAYT CFGTPANYPL FAGVWAAVLA
451 GCILRHRKDL HKLFHYLKKQ GFPL*

Further sequence analysis revealed the complete DNA sequence<SEQ ID 375> to be:

25
30
35
40
45
50
1 ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC
51 GGTTTTAGCC GTCATCATCC TGGCGCTGCT GTCGTGGTAT TTCCCGCGCG
101 ACGACATCGG GCGCATCGTG CTGATGCAGA CGGCGGCGGG GCTGACGGTG
151 TCGGTGTTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
201 CACCGCCGAC AAAGACACCT TGTTCAAAAC CCTGTTCTCG CCGCCGCTGC
251 TGTCTGCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
301 TCTGAAATCC TGTTCCTACT CGACGATGCC GCCGCCGGCA TCGGGCTGGT
351 GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTGC
401 GTATGGAAGG ACGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGCCCAAG
451 CTCGCCATCC TGCTGCTGCT GCCGCTGACG GTCGGGCTGC TGCACTTTCC
501 AGCGAACACC GCCGTCTGA CCGCGTTTA CCGCTGGCA AACCTTGCCG
551 CCGCCGCCCT TTTGCTGTTT CAAAACCGAT GCCGTCTGAA GGCCGTCCGG
601 CACGCACCGT TTTCCGCCGC CGTCCTGCAC CGGGGGCTGC GCTACGGCAT
15 651 ACCGATCGCA CTGAGCAGCA TCGCTATTG GGGGCTGGCA TCCGCCGACC
701 GTTTGTTCCT GAAAAAATAT GCCGGCCTGG AACAGCTCGG CGTTTATTCG
751 ATGGGTATTT CGTTCGGCGG GCGGCATTA TTGTTCCAAA GCATCTTTTC
801 AACGGTCTGG ACACCGTATA TTTCCGCGC AATCGAAGAA AACCCCCGC
851 CCGCCCGCCT CTCGCAACG GCAGAATCCG CCGCCGCCCT GCTGCTCTCC
40 901 GCCCTCTGCC TGACCGGCAT TTTCTCGCCC CTTGCTCTCC TCCTGTGCC
951 GGAAACTAC GCCCGCGTCC GGTTCATCGT CGTATCGTG ATGCTGCCGC
1001 CGCTGTTTTC CACGCTGGCG GAAATCAGCG GCATCGGTTT GAACGTCGTC
1051 CGCAAAACGC GCCCGATCGC GCTCGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CCTGCTGCTG CTGGGGCTTG CCGTGCCGTC CGGCGGCGCG CGCGGCGCGG
45 1151 CGGTTCCTGT TGCCGCTCA TTTCTGGCTGT TTTTTCCTT CAAGACCGAA
1201 AGCTCCTGCC GCCTGTGGCA GCCGCTCAA CGCCTGCCG TTTATCTGCA
1251 CACATTGTTC TGCTGACCT CCTCGCGGCG CTACACCTGC TTCGGCACGC
1301 CGGCAAACTA TCCCTGTTT GCCGCGGTAT GGGCGGCATA TCTGGCAGGC
1351 TGCATCCTGC GCCACCGGAA AGATTGAC AAATGTTTC ATTATTGAA
50 1401 AAAACAAGGT TTCCATTAT GA

This corresponds to the amino acid sequence <SEQ ID 376; ORF10-1>:

55
60
1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YVREYYATAD KDTLFKTLFL PPLLSAAAIA ALLLSRPSLP
101 SEILFSLDDA AAGIGLVLE LSLPIRELL LVLRMEGRAL AFSSAQLVPK
151 LAIIILLXPLT VGLLHFPANT AVLTAVYALA NLAAAFLLF QNRCRLKAVR
201 HAPFSPAVLH RGLRYGIPIA LSSIAYWGLA SADRLFLKKY AGLEQLGVYS
251 MGISFGGAAL LFQSIYSTVW TPYIFRAIEE NAPPARLSAT AESAAALLAS
301 ALCLTGIFSP LASLLLPENY AAVRFIVVSC MLPLFCTLA EISGIGLNVV
351 RKTRPIALAT LGALANLLL LGLAVPSGGA RGAAVACAAS FWLFFAFKTE
401 SSCRLWQPLK RLPYLYLHTL CLTSSAAYTC FGTPANYPLF AGVWAAVLAG
451 CILRHRKDL KLFHYLKKQGFPL*

Computer analysis of this amino acid sequence gave the following results:

Prediction

ORF10-1 is predicted to be the precursor of an integral membrane protein, since it comprises several (12-13) potential transmembrane segments, and a probable cleavable signal peptide

Homology with EpsM from *Streptococcus thermophilus* (accession number U40830).

- 5 ORF10 shows homology with the *epsM* gene of *S. thermophilus*, which encodes a protein of a size similar to ORF10 and is involved in expolysaccharide synthesis. Other homologies are with prokaryotic membrane proteins:

Identities = (25%)

10 Query: 213 LRYGIPLALSSLAYWGLASADRLFLKKYAGLEQLGVYSMGISFGGAALLQSIFSTVW 270
 L Y +PL SS+ +W L ++ R F+ + G G+ ++ + +IF+ W
 Sbjct: 210 LYYALPLIPSSILWLLNASSRYFVLFLLGAGANGLAVATKIPSIISIFNTIFTQAW 267

15 Identities = 15/57 (26%), Positives = 31/57 (54%)

Query: 7 LGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQAYVR 63
 L + G++GS +L +++PL ++ + G L Q T A L + ++ + + A +R
 Sbjct: 12 LVFTIGNLGSKLLVFLVPLYTYAMTPQEYGMADLYQTTANLLLPLITMNVFDTLR 68

20 Identities = 16/96 (16%), Positives = 36/96 (37%)

Query: 307 IFSPLASLLLPENYAAVRFTVVSCLPPLFYTLTEISGIGLVVRKTRPIXXXXXXXXXX 366
 + P+ ++ +YA+ V ML LF + ++ G ++T+ +
 25 Sbjct: 305 VLKPIVEKVVSSDYASSQYVFFMLSMFLSSFSDFGTNYIAAKQTKGVFMTSIYGTIV 364

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF10 shows 95.4% identity over a 475aa overlap with an ORF (ORF10a) from strain A of *N.*

meningitidis:

30	orf10.pep	MDTKEILXYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA
	orf10a	MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA
35	orf10.pep	YVREYYATADKDTLFTLFLPPLLSAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE
	orf10a	YVREYYAAADKDTLFTLFLPPLLSAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE
40	orf10.pep	LSFLPIRIFLLLVLRMEGRALAFSSAQLVSKLAILLLXPLTVGLLHFPANTAVLTAVYALA
	orf10a	LSFLPIRIFLLLVLRMEGRALAFSSAQLVSKLAILLLXPLTVGLLHFPANTAVLTAVYALA
50	orf10.pep	NLAAAFLFQNRCLKAVRHAPFSPAVLHRGXRYGIPIALSSIAWGLASADRLFLKKY
	orf10a	NLAAAFLFQNRCLKAVRRAPFSSAVLHRGLRYGIPIALSSIAWGLASADRLFLKKY
55	orf10.pep	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEENAPPARLSATAESAAALLAS
	orf10a	AGLEQLGVYSMGISFGGAALLFQSIFSTVWTPYIFRAIEANAPPARLSATAESAAALLAS

		310	320	330	340	350	360
	orf10.pep	ALCXTGIFSP	LASLLPEN	YAAVRFIV	VSCMXPPL	FCTLAEIS	GIGLNVVRKTRPIALAT
5	orf10a	ALCLTGIFSP	LASLLPEN	YAAVRFIV	VSCMLPPL	FCTLVEIS	GIGLNVVRKTRPIALAT
		310	320	330	340	350	360
	orf10.pep	LGALAANLLL	LGLDRAVP	PAR-PXGA	AVACAASF	WLFFAFKTE	SSCRLWQPLKRLPLYLHT
10	orf10a	LGALAANLLL	LGL--AVP	SGGARGAA	VACAASF	WLFFVFKTE	SSCRLWQPLKRLPLYMHT
		370	380	390	400	410	419
	orf10.pep	LFCLTSSAAY	TCFGTPAN	YPLFAGVW	AAAYLAGC	ILRHRKDL	HLKLFHYLKKQGFPLX
15	orf10a	LFCLASSAAY	TCFGTPAN	YPLFAGVW	AVYLAGC	ILRHRKDL	HLKLFHYLKKQGFPLX
		420	430	440	450	460	470

The complete length ORF10a nucleotide sequence <SEQ ID 377> is:

20	1	ATGGACACAA	AAGAAATCCT	CGGCTACGCG	GCAGGCTCGA	TCGGCAGCGC
	51	GGTTTTAGCG	GTCAATCATCC	TGCCGCTGCT	GTCTGTGGTAT	TTCCCTGCCG
	101	ACGACATCGG	ACGCATCGTG	CTGATGCAGA	CGGCGGCGGG	GCTGACGGTG
	151	TCGGTGTGTG	GCCTCGGGCT	GGATCAGGCA	TACGTCCGCG	AATACTATGC
	201	CGCCGCCGAC	AAAGACACTT	TGTTCAAAAC	CCTGTTCTCTG	CCGCCGCTGC
25	251	TGCTGCGCG	CGCGATAGCC	GCCCTGCTGC	TTTCCCGCCC	ATCCCTGCCG
	301	TCTGAAATCC	TGTTTTCGCT	CGACGATGCC	GCCGCCGGCA	TCGGGCTGGT
	351	GCTGTTTGAA	CTGAGCTTCC	TGCCCATCCG	CTTCTCTTA	CTGGTTTTCG
	401	GTATGGAAG	ACGCGCCCTT	GCCTTTTCGT	CCGCGCAACT	CGTGTCCAAG
	451	CTCGCCATCC	TGCTGCTGCT	GCCGCTGACG	GTCCGGCTGC	TGCACTTTCC
30	501	GGCGAACACC	GCCGTCCTGA	CCGCGTTTA	CGCGCTGGCA	AACCTTGCCG
	551	CCGCCGCCTT	TTTGCTGTTT	CAAAACCGAT	GCCGCTCGAA	GGCCGTCCGG
	601	CGCGCACCGT	TTTCATCCGC	CGTCTGCAT	CGCGGCTGCG	GCTACGGCAT
	651	ACCGATCGCA	CTAAGCAGCA	TCGCCTATTG	GGGGCTGGCA	TCCGCCGACC
	701	GTTTGTTCCT	GAAAAAATAT	GCCGGCCTAG	AACAGCTCGG	CGTTTATTCG
35	751	ATGGGTATTT	CGTTCGGCGG	AGCGGCATTA	TTGTTCAAA	GCATCTTTTC
	801	AACGGTCTGG	ACACCGCTATA	TTTTCCGCGC	AATCGAAGCA	AACGCCCCCG
	851	CCGCCCGCCT	CTCGGCAACG	GCAGAATCCG	CCGCCGCCCT	GCTTGCCCTC
	901	GCCCTCTGCC	TGACCGGCAT	TTTCTCGCCC	CTCGCTCCC	TCCTGCTGCC
	951	GGAAACTAC	GCCGCCGTCC	GGTTATCGT	CGTATCGTGT	ATGCTGCCTC
40	1001	CGCTGTTTTG	CACGCTGGTA	GAAATCAGCG	GCATCGGTTT	GAACGTGCTC
	1051	CGAAAAACAC	GCCCGATCGC	GCTCGCCACC	TTGGGCGCGC	TGGCGGCAAA
	1101	CCTGCTGCTG	CTGGGGCTTG	CCGTACCGTC	CGGCGGCGCG	CGCGGCGCGG
	1151	CGGTTGCCTG	TGCCGCCTCA	TTTTGGCTGT	TTTTTGT	CAAGACCGAA
	1201	AGCTCTGCGC	GCCTGTGGCA	GCCGCTCAAA	CGCCTGCCGC	TTTATATGCA
45	1251	CACATTGTTC	TGCCTGGCCT	CCTCGGCGGC	CTACACTGCG	TTCGGCACTC
	1301	CGGCAAACTA	CCCCCTGTTT	GCCGGCGTAT	GGGCGGTATA	TCTGGCAGGC
	1351	TGCATCCTGC	GCCACCGGAA	AGATTGCAC	AAACTGTTTC	ATTATTGAA
	1401	AAAACAAGGT	TTCCATTAT	GA		

This encodes a protein having amino acid sequence <SEQ ID 378>:

50	1	MDTKEILGYA	AGSIGSAVLA	VIILPLLSWY	FPADDIGRIV	LMQTAAGLTV
	51	SVLCLGLDQA	YVREYAAAD	KDTLFKTLFL	PPLLSAAAIA	ALLLSRPSLP
	101	SEILFSLDDA	AAGIGLVLFE	LSFLPIRFL	LVLRMGRAL	AFSSAQLVSK
	151	LAILLPLT	VGLLHFPANT	AVLTAVYALA	NLAAAFLLF	QNRCLKAVR
	201	RAPFSSAVLH	RGLRYGIPIA	LSSIAYWGLA	SADRLFLKKY	AGLEQLGVYS
55	251	MGISFGGAAL	LFQSFSTVW	TPYIFRAIEA	NAPPARLSAT	AESAAALLAS
	301	ALCLTGIFSP	LASLLPENY	AAVRFIVVSC	MLPPLFCTLV	EISGIGLNVV
	351	RKTRPIALAT	LGALAANLLL	LGLAVPSGGA	RGAAVACAAS	FWLFFVEKTE
	401	SSCRLWQPLK	RLPLYMHTLF	CLASSAAYTC	FGTPANYPLF	AGVWAVYLAG
	451	CILRHRKDLH	KLFHYLKKQG	FPL*		

ORF10a and ORF10-1 show 95.4% identity in 475 aa overlap:

60		10	20	30	40	50	60
	orf10-1.pep	MDTKEILXYA	AGSIGSAVL	AVIILPLLSWY	FPADDIGRIV	LMQTAAGLTV	SVLCLGLDQA
	orf10a	MDTKEILGYA	AGSIGSAVL	AVIILPLLSWY	FPADDIGRIV	LMQTAAGLTV	SVLCLGLDQA
65		10	20	30	40	50	60

-240-

		70	80	90	100	110	120
	orf10-1.pep	YVREYYATADKDTL	FKTLFLPPLLSAA	IAALLSRPSLPSE	ILFSLDDAAAG	IGLVLFE	
5	orf10a	YVREYYAAADKDTL	FKTLFLPPLLSAA	IAALLSRPSLPSE	ILFSLDDAAAG	IGLVLFE	
		70	80	90	100	110	120
	orf10-1.pep	LSFLPIRFLLLVLR	MEGRALAFSSAQ	LVPKLAILLXPL	TVGLLHFPANT	AVLTAVYALA	
10	orf10a	LSFLPIRFLLLVLR	MEGRALAFSSAQ	LVPKLAILLXPL	TVGLLHFPANT	AVLTAVYALA	
		130	140	150	160	170	180
	orf10-1.pep	NLAAAFLFQNR	CRLKAVRHAPF	SPAVLHRGXRY	GIPIALSSIA	YWGLASADRL	FLKKY
15	orf10a	NLAAAFLFQNR	CRLKAVRRAPF	SAVLHRGLRY	GIPIALSSIA	YWGLASADRL	FLKKY
		190	200	210	220	230	240
	orf10-1.pep	AGLEQLGVYS	MGISFGGAALL	FQSIFSTVWTP	YIFRAIEEN	APPARLSATA	ESAALLAS
20	orf10a	AGLEQLGVYS	MGISFGGAALL	FQSIFSTVWTP	YIFRAIEAN	APPARLSATA	ESAALLAS
		250	260	270	280	290	300
	orf10-1.pep	ALCXTGIFSP	PLASLLLPEN	YAAVRFIVV	SCMPPLFCT	LAEISGIGL	NNVVRKTRPIALAT
25	orf10a	ALCLTGIFSP	PLASLLLPEN	YAAVRFIVV	SCMLPPLF	CTLVEISGIGL	NNVVRKTRPIALAT
		310	320	330	340	350	360
	orf10-1.pep	LGALAAANLLL	GLDRAVPA	R-PXGA	AVACAASFWL	FFAFKTESS	CRWLQPLKRLPLYLHT
30	orf10a	LGALAAANLLL	GLL--AV	PSGGARGA	AVACAASFWL	FFVFKTESS	CRWLQPLKRLPLYMHT
		370	380	390	400	410	419
	orf10-1.pep	LFCLTSSA	AYTCFGTP	PANYPLFAG	VWAAAYLAG	CILRHRKDL	HKLHFHYLKKQGFPLX
35	orf10a	LFCLASSA	AYTCFGTP	PANYPLFAG	VWAVYLAG	CILRHRKDL	HKLHFHYLKKQGFPLX
		420	430	440	450	460	470
	orf10-1.pep	MDTKEIL	GYAAGSIG	SAVLAVI	ILPPLLSWY	FPADDIGR	IVLMQTAAGLTVSVLCLGLDQA
40	orf10a	MDTKEIL	XYAAGSIG	SAVLAVI	ILPPLLSWY	FPADDIGR	IVLMQTAAGLTVSVLCLGLDQA

Homology with a predicted ORF from *N.gonorrhoeae*

ORF10 shows 94.1% identity over a 475aa overlap with a predicted ORF (ORF10.ng) from *N.*

45 *gonorrhoeae*:

	orf10ng.pep	MDTKEILGYAAGSIGSAVLAVIILPPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	60
	orf10nm	MDTKEILXYAAGSIGSAVLAVIILPPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA	60
50	orf10ng.pep	YVREYYAAADKDTLFKTLFLPPLLSAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE	120
	orf10nm	YVREYYATADKDTLFKTLFLPPLLSAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE	120
	orf10ng.pep	LSFLPIRFLLLVLRMEGRALAFSSAQ LVPKLAILLXPLTVGLLHFPANTS SVLTAVYALA	180
55	orf10nm	LSFLPIRFLLLVLRMEGRALAFSSAQ LVPKLAILLXPLTVGLLHFPANTAVLTAVYALA	180
	orf10ng.pep	NLAAAFLFQNR CRLKAVRRAPFSPAVLHRGLRYGIPIALSSLAYWGLASADRLFLKKY	240
60	orf10nm	NLAAAFLFQNR CRLKAVRHAPFSPAVLHRGXRYGIPIALSSIA YWGLASADRLFLKKY	240
	orf10ng.pep	AGLEQLGVYS MGISFGGAALLQSIFSTVWTPYIFRAIEENATPARLSATAESAALLAS	300
	orf10nm	AGLEQLGVYS MGISFGGAALLQSIFSTVWTPYIFRAIEENAPPARLSATAESAALLAS	300
65	orf10ng.pep	ALCLTGIFSP LASLLLPENYAAVRFTVVS CMLPPLFYTLTEISGIGLNNVVRKTRPIALAT	360
	orf10nm	ALCXTGIFSP LASLLLPENYAAVRFIVVSCMPPLFCTLA EISGIGLNNVVRKTRPIALAT	360

-241-

```

      370      380      390      400      410
orfl0ng.pep LGALAANLLLLGL--AVPSGGTRGA AVACAASFWLFFVFKTESSCRLWQPLKRLPLYMHT
5 orfl0nm    LGALAANLLLLGLDRAVPA R-PXGAAVACAASFWLFFAFKTESSCRLWQPLKRLPLYLHT
      370      380      390      400      410

      420      430      440      450      460      470
10 orfl0ng.pep LFCCLASSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKNLHKLHFHYLKKQGFPPLX
orfl0nm    LFCLTSSAAYTCFGTPANYPLFAGVWAAYLAGCILRHRKDLHKLHFHYLKKQGFPPLX
      420      430      440      450      460      470

```

The complete length ORF10ng nucleotide sequence <SEQ ID 379> is:

```

15 1 ATGGACACAA AAGAAATCCT CGGCTACGCG GCAGGCTCGA TCGGCAGCGC
51 GGTTTTAGCC GTCATCATCC TGCCGCTGCT GTCGTGGTAT TTCcccgCCG
101 ACGACATCGG GCGCATCGTG CTGATGCAGA CGGCGCGGGG ACTGACGGTG
151 TCGGTATTGT GCCTCGGGCT GGATCAGGCA TACGTCCGCG AATACTATGC
201 CGCCGCCGAC AAAGACACTT TGTTCAAAAC CCTGTTCTCTG CCGCCGCTGC
251 TGTTTTCCGC CGCGATAGCC GCCCTGCTGC TTTCCCGCCC GTCCCTGCCG
20 301 TCTGAAATCC TGTTTTGCTT CGACGATGCC GCCCGCGGCA TCGGGCTGGT
351 GCTGTTTGAA CTGAGCTTCC TGCCCATCCG CTTTCTCTTA CTGGTTTTCG
401 GTATGGAAGG GCGCGCCCTT GCCTTTTCGT CCGCGCAACT CGTGCCCAAA
451 CTCGCCATTC TGCTGCTGTT GCCGCTGACG GTCGGGCTGC TGCACTTTCC
501 GCGGAACACC TCCGTCCTGA CCGCCGTTTA CGCGCTGGCA AACCTTGCCG
25 551 CCGCCGCCCT TTTGCTGTTT CAAACCGAT GCCGTCTGAA GGCCGTCCGG
601 CGCGCGCCGT TTTGCGCCGC CGTCTGCAC CGGGGCTGCG GCTACGGCAT
651 ACGGCTCGCA CTGAGCAGCC TTGCCTATTG GGGGCTGGCA TCCGCGGACC
701 GTTTGTTCTT GAAAAAATAT GCGGGCCTGG AACAGCTCGG CGTTTATTTC
751 ATGGGTATTT CGTTCGCGCG GCGCGCATTA TTGCTCCAAA GCATCTTTTC
30 801 AACGGTCTGG ACACCGTATA TTTCCGTGCG AATCGAAGAA AACGCCACGC
851 CGGCCCGCCT CTCGGCAACG GCAGAATCCG CCGCCGCCCT GCTTGCTTCC
901 GCCCTCTGCC TGACCGGAAT TTTCTCGCCC CTCGCTCCC TCCTGTGCC
951 GGAAACTAC GCGCGCTCC GGTTTACCGT CGTATCGTGT ATGCTGccgc
1001 cgctGTTTTA CACGCTGACC GAAATCAGCG GCATCGGTTT GAACGTCGTC
35 1051 CGCAAAACGC GTCCGATCGC GCTTGCCACC TTGGGCGCGC TGGCGGCAAA
1101 CCTGCTGCTG CTGGGGCTTG CCGTACCGTC CGGCGGCACG CGCGGCGCGG
1151 CGGTTGCCTG TGCCGCCTCA TTCTGGTTGT TTTTGTTTT CAAGACAGAA
1201 AGCTCCTGCC GCCTGTGGCA GCCGCTCAA CGCCTGCCG TTTATATGCA
1251 CACATTGTTT TGCTTgCCT CTTGCGCGGC CTACACCTGC TTCGGCACAC
40 1301 CGGCAACTA CCCcctgttt gccggcgtAT GGGCGGCATA TCTGGCAGGC
1351 TGCATCCTGC GCCACCGGAA AAATTTGCAC AAAGTGTTC ATTATTGAA
1401 AAAACAAGGT TTCCATTAT GA

```

This encodes a protein having amino acid sequence <SEQ ID 380>:

```

45 1 MDTKEILGYA AGSIGSAVLA VIILPLLSWY FPADDIGRIV LMQTAAGLTV
51 SVLCLGLDQA YVREYYAAAD KDTLFKTLFL PPLLFSAAIA ALLLSRPSLP
101 SEILFSLDDA AAGIGLVLFE LSFLPIRFL LVLMEGRAL AFSSAQLVFK
151 LAIIIIIIPLT VGLLHFPANT SVLTAVYALA NLAAAFLLF QNRCRLKAVR
201 RAPFSPAVLH RGLRYGIPLA LSSLAYWGLA SADRLFLKKY AGLEQLGVYS
50 251 MGISFGGAAL LLQSIFSTVW TPYIFRAIEE NATPARLSAT AESAAALLAS
301 ALCLTGIFSP LASLLLPENY AAVRFTVWSC MLPPLFYTLT EISGIGLNVV
351 RKTRPIALAT LGALAANLL LGLAVPSGGT RGA AVACAAS FWLFFVFKTE
401 SSCRLWQPLK RLPLYMHTLF CLASSAAYTC FGTPANYPLF AGVWAAYLAG
451 CILRHRKNLH KLFHYLKKQG FPL*

```

ORF10ng and ORF10-1 show 96.4% identity in 473 aa overlap:

```

55 10 20 30 40 50 60
orfl0-1.pep MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA
orfl0ng-1 MDTKEILGYAAGSIGSAVLAVIILPLLSWYFPADDIGRIVLMQTAAGLTVSVLCLGLDQA
60 10 20 30 40 50 60
70 80 90 100 110 120
orfl0-1.pep YVREYYATADKDTLFKTLFLPPLLSSAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE
orfl0ng-1 YVREYYAAADKDTLFKTLFLPPLLSSAAIAALLSRPSLPSEILFSLDDAAAGIGLVLFE
65 70 80 90 100 110 120

```


Based on this analysis, including the presence of a putative leader peptide and several transmembrane segments and the presence of a leucine-zipper motif (4 Leu residues spaced by 6 aa, shown in bold), it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 381>:

This corresponds to the amino acid sequence <SEQ ID 382; ORF65>:

60 1 . . I L K P H N Q L K E D I O P D P A D Q N A L S E P D A A T E A E Q S D A E N A A D K Q P V A D K A D
51 E V E E K A G E P E R E E P D G Q A V R K K A L T E E R E Q T V R E K A Q K K D A E T V K I Q A V K

Further work revealed the complete nucleotide sequence <SEQ ID 383>:

This corresponds to the amino acid sequence <SEQ ID 384; ORF65-1>:

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF65 shows 92.0% identity over a 150aa overlap with an ORF (ORF65a) from strain A of *N. meningitidis*:

[illegible]

The complete length ORF65a nucleotide sequence <SEQ ID 385> is:

1 ATGTTTATGA ACAAATTTTC CCAATCCGGA AAAGGTCTGT CCGGTTTTTT
51 CTTCGGTTTG ATACTGCCGA CGGTCATTAT TGCCGGTATT TTGTTTTATC

101 TGAACCAGAG CGGTCAAAAT GCGTTCAAAA TCCCGGTTCC GTCGAAGCAG
 151 CCTGCAGAAA CGGAAATCCT GAAACCGAAA AACCAGCCTA AGGAAGACAT
 201 CCAACCTGAA CCGGCCGATC AAAACGCCTT GTCCGAACCG GATGCTGCGA
 251 AAGAGGCAGA GCACTCGGAT GCGGAAAAAG CTGCCGACAA GCAGCCCGTT
 301 GCCGACAAAG CCGACGAGGT TGAGGAAAAAG GCGGACGAGC CGGAGCGGGA
 351 AAAGTCGGAC GGCAGGCAG TGCGCAAGAA AGCACTGACG GAAGAGCGTG
 401 AACAAACCGT CGGGGAAAAA GCGCAGAAGA AAGATGCCGA AACCGTTAAA
 451 AAACAAGCGG TAAAACCATC TAAAGAAACA GAGAAAAAAG CTTCAAAGA
 501 AGAGAAAAAG GCGGAGAAGG AAAAAGTTGC ACCCAAACCG ACCCGGAAC
 551 AAATCCTCAA CAGCGGCAGC ATCGAAAAAG CGCGCAGTGC CGCTGCCAAA
 601 GAAGTGCAGA AAATGAAAAA GCCCGACAAG GCGGAAGCAA CGCATTATCT
 651 GCAAATGGGC GCGTATGCCG ACCGCCGAG CGCGGAAGGG CAGCGTGCCA
 701 AACTGGCAAT CTTGGGCATA TCTTCCAAGG TGGTCGGTTA TCAGGCGGGA
 751 CATAAACGCG TTTACCGGGT GCAAAGCGGC AATATGTCTG CCGATGCGGT
 801 GAAAAAATG CAGGACGAGT TGA AAAAACA TGAAGTCGCC AGCCTGATCC
 851 GTTCTATCGA AAGCAAATAA

This encodes a protein having amino acid sequence <SEQ ID 386>:

1 MFMNKSQSG KGLSGFFFL ILATVIIAGI LFYLNQSGON AFKIPVPSKQ
 51 PAETEILKPK NQPKEDIQPE PADQNALSEP DAAKEAEQSD AEKAADKQPV
 101 ADKADEVEEK ADEPEREKSD QAVRKKALT EEREQTVGEK AQKKDAETVK
 151 KQAVKPSKET EKKASKEEK AEKEKVAPKP TPEQILNSGS IEKARSAALK
 201 EVQKMKTPDK AEATHYLQMG AYADRRSAEG QRAKLAILGI SSKVVGQYQAG
 251 HKTLYRVQSG NMSADAVKKM QDELKKHEVA SLIRSIESK*

ORF65a and ORF65-1 show 96.5% identity in 289 aa overlap:

25	orf65a.pep	10	20	30	40	50	60
	orf65-1	10	20	30	40	50	60
30	orf65a.pep	70	80	90	100	110	120
	orf65-1	70	80	90	100	110	120
35	orf65a.pep	130	140	150	160	170	180
	orf65-1	130	140	150	160	170	180
40	orf65a.pep	190	200	210	220	230	240
	orf65-1	190	200	210	220	230	240
45	orf65a.pep	250	260	270	280	290	
	orf65-1	250	260	270	280	290	

55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF65 shows 89.6% identity over a 212aa overlap with a predicted ORF (ORF65.ng) from *N. gonorrhoeae*:

60	ORF65ng	30	40	50	60	70	80
	ORF65	10	20	30			

-245-

		90	100	110	120	130	140
ORF65ng	AEQSDAEKAADKQPVADKADVEEEKAGEPEREEDGQAVRKKALTEEREQTVREKAQKKD						
5	ORF65	AEQSDAENAADKQPVADKADVEEEKAGEPEREEDGQAVRKKALTEEREQTVREKAQKKD	40	50	60	70	80
		150	160	170	180	190	200
ORF65ng	AETVKKKAVKPSKETEKKASKEEKAAKEKVAPKPTPEQILNSRSIEKARSAAAKEVQKM						
10	ORF65	AETVKIQAVKPSKETEKKASKEEKAAKEKVAPKPTPEQILNSGSIEXARSAAAKEVQKM	100	110	120	130	140
		210	220	230	240	250	260
ORF65ng	KNFGQGGSQRIICKWARMNPNGARKGSPVNPQSWAYLPKWSAIRRDIKRFTACKAAICPP						
15	ORF65	XNVRQGGSXRIICKWARMPTVRARKGSPVNPQSWAYLPKWSAIRRDIKRFTGCKAAICLP	160	170	180	190	200
		210	220	230	240	250	260
ORF65ng	MR						
20	ORF65	MR					

An ORF65ng nucleotide sequence <SEQ ID 387> was predicted to encode a protein having amino acid sequence <SEQ ID 388>:

1	MF MNKFSQSG KGLSGFFFL ILATVIIAGI LLYLNQGGON AFKIPAPSKQ
51	PAETEILKLK NQPKEDIQPE PADQNALSEP DVAKEAEQSD AEKAADKQPV
101	ADKADEVEEK AGEPEREEDP GQAVRKKALT EEREQTVREK AQKKDAETVK
151	KKAVKPSKET EKKASKEEK AAKEKVAPKP TPEQILNSRS IEKARSAAAK
201	EVQKMKNFGQ GGSQRIICKW ARMPNPGARK GSVNPQSWA YLPKWSAIRR
251	DIKRFTACKA AICPPMR*

After further analysis, the complete gonococcal DNA sequence <SEQ ID 389> was found to be:

1	ATGTTTATGA ACAAATTTTC CCAATCCGGA AAAGGTCTGT CCGGTTTCTT
51	CTTCGGTTTG ATACTGGCAA CGGTTCATTAT TGCCGGTATT TTGCTTTATC
101	TGAACCAGGG CGGTCAAAT GCGTTCAAAA TCCCGGCTCC GTCGAAGCAG
151	CCTGCAGAAA CGGAAATCCT GAAACTGAAA AACCAGCCTA AGGAAGACAT
201	CCAACCTGAA CCGGCCGATC AAAACGCCTT GTCCGAACCG GATGTTGCGA
251	AAGAGGCAGA GCAGTCGGAT GCGGAAAAAG CTGCCGACAA GCAGCCCGTT
301	GCCGACAAag ccgacgAGGT TGAAGAAAag GcGGgagAgc cggAACGGga
351	aGAGCCGGAC ggACAGGCAG TCGCAAGAA AGCACTGAcg gAAGAgcGTG
401	AACAAACcgt cagggAAAAA GCGCagaaga AAGATGCCGA AACGgTTAAA
451	AAacaaGCg tAaaaccgtc tAAAGAAACa gagaaaaaag cTcaaaaaa
501	agagaaaaag gcggcgaaa gaaaAGttgc acccaaaccg accccggaaC
551	aaatcctcaa cagccgCagc atcgaaaaag cgcgtagtgc cgtgcca
601	gaAgtgcaGA AAatgaaaaa ctTtgggcaa ggcgGaagcc aacgcattaT
651	CTGcaaatgg gcgcgtatgc cgaccgtccg gagcgcggaA gggcagcgtg
701	ccaaACtggc aAtcttgGgc atatctTccg aagtggctcg CTATCAGGCG
751	GGACATAAAA CGCTTTACCG CGTGCAAagc GGCAaatatgt ccgccgatgc
801	gGTGAAAAA ATGCAGGACG AGTTGAAAA GCATGGGGtt gcCAGCCTGA
851	TCCGTGcgAT TGAAGGCAAA TAA

This encodes the following amino acid sequence <SEQ ID 390>:

1	MF MNKFSQSG KGLSGFFFL ILATVIIAGI LLYLNQGGON AFKIPAPSKQ
51	PAETEILKLK NQPKEDIQPE PADQNALSEP DVAKEAEQSD AEKAADKQPV
101	ADKADEVEEK AGEPEREEDP GQAVRKKALT EEREQTVREK AQKKDAETVK
151	KQAVKPSKET EKKASKEEK AAKEKVAPKP TPEQILNSRS IEKARSAAAK
201	EVQKMKNFGQ GGSQRIICKW ARMPTVRSAG GQRAKLAILG ISSEVVGYQA
251	GHKTLYRVQS GNMSADAVKK MQDELKKHGV ASLIRAIEGK *

ORF65ng-1 and ORF65-1 show 89.0% identity in 290 aa overlap:

		10	20	30	40	50	60
orf65-1.pep	MF MNKFSQSGKGLSGFFFLILATVIIAGILFYLNQSGQNAFKIPASSKQPAETEILKPK						
60	orf65ng-1	MF MNKFSQSGKGLSGFFFLILATVIIAGILLYLNQGGQNAFKIPAPSKQPAETEILKLK	10	20	30	40	50

-246-

		70	80	90	100	110	120
	orf65-1.pep	NQPKEDIQPEPADQNALSEPDAATEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEDP					
5	orf65ng-1	NQPKEDIQPEPADQNALSEPDAATEAEQSDAEKAADKQPVADKADEVEEKAGEPEREEDP					
		70	80	90	100	110	120
	orf65-1.pep	GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETEKKASKEEKKAKEKVAPKP					
10	orf65ng-1	GQAVRKKALTEEREQTVREKAQKKDAETVKKQAVKPSKETEKKASKEEKKAKEKVAPKP					
		130	140	150	160	170	180
	orf65-1.pep	TPEQILNSGSIEKARSAAAKEVQKMKTSDKAEATHYL-OMGAYADRQSAEQRAKLAILG					
15	orf65ng-1	TPEQILNSRSIEKARSAAAKEVQKMKNFQGGSQRIICKWARMPTVRSAEQRAKLAILG					
		190	200	210	220	230	239
	orf65-1.pep	ISSKVVGYYAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX					
20	orf65ng-1	ISSEVVGYYAGHKTLYRVQSGNMSADAVKKMQDELKKHGVASLIRAIEGKX					
		240	250	260	270	280	290
	orf65-1.pep	ISSKVVGYYAGHKTLYRVQSGNMSADAVKKMQDELKKHEVASLIRSIESKX					
	orf65ng-1	ISSEVVGYYAGHKTLYRVQSGNMSADAVKKMQDELKKHGVASLIRAIEGKX					
		250	260	270	280	290	

- 25 On this basis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 46

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

30 391>:

	1	ATGAACCACG	ACATCACTTT	CCTCACCCCTG	TTCCTACTCG	GTCTCTTCGG
	51	CGGAACGCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GcGTTTGs.s
	101	TCCAACCTCC	CCCGCATATC	AACCGCTTTT	GGCTGATCCT	GCTGCTTAAC
	151	ACAGGACGGG	TAAGCAGCTA	TACGGCAATC	GGCCTGATAC	TCGGATTAAT
35	201	CGGACAGGTC	GGCGTTTCAC	TCGACCAaAC	CCGCGTCCTG	CAGAATATTT
	251	TATACACGGC	CGCCAACCTC	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
	301	GGTATTTCCT	CCTTGGCGGC	AAAAATCGAG	AAAATCGGCA	AACCGATATG
	351	GCGGAACCTG	AACCCGATAC	TCAACCGGCT	GTTACCCATA	AAATCCATAC
	401	CCGCCTGCCT	tCGGtTCGGA	ATATTATGGG	GCTGGCTGCC	GTGCGGACTG
40	451	GTTTACAGCG	CGTCGCTTTA	CGCGCTGGGA	AgCGGTAGTG	CGGCAACGGG
	501	CGGGTTATAT	ATGCTTGCCT	TTGCACTGGG	TACGCTGCCC	AATCTTtTAG
	551	CAATCGGCAT	TTTtTCCCTG	CAACTGAawa	AAATCATGCA	AAACCGATAT
	601	ATCCGCCTGT	GTACGGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAACT
	651	TGCCGTCCTG	TGGCTGTAA			

- 45 This corresponds to the amino acid sequence <SEQ ID 392; ORF103>:

	1	MNHDITFTL	FLLGXFGGTH	CIGMCGGLSS	AFXXQLPPhi	NRFWLILLN
	51	TGRVSSYTAI	GLILGLIGQV	GVSLDQTRVL	QNILYTAANL	LLLFLGLYLS
	101	GISSLAAKIE	KIGKPIWRNL	NPILNRLLPI	KSIPACLAVG	ILWGWLPcGL
	151	VYSASLYALG	SGSAATGGLY	MLAFALGTLF	NLLAIGIFSL	QLXKIMQnRY
50	201	IRLCTGLSVS	LWALWKLAVL	WL*		

Further work elaborated the DNA sequence <SEQ ID 393> as:

	1	ATGAACCACG	ACATCACTTT	CCTCACCCCTG	TTCCTACTCG	GTTTCTTCGG
	51	CGGAACGCAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GCGTTTGC
	101	TCCAACCTCC	CCCGCATATC	AACCGCTTTT	GGCTGATCCT	GCTGCTTAAC
	151	ACAGGACGGG	TAAGCAGCTA	TACGGCAATC	GGCCTGATAC	TCGGATTAAT
55	201	CGGACAGGTC	GGCGTTTCAC	TCGACCAaAC	CCGCGTCCTG	CAGAATATTT
	251	TATACACGGC	CGCCAACCTC	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
	301	GGTATTTCCT	CCTTGGCGGC	AAAAATCGAG	AAAATCGGCA	AACCGATATG

-247-

351 GCGGAACCTG AACCCGATAC TCAACCGGCT GTTACCCATA AAATCCATAC
 401 CCGCCTGCCT TCGCGTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTG
 451 GTTTACAGCG CGTCGCTTTA CGCGCTGGGA AGCGGTAGTG CGGCAACGGG
 501 CGGGTTATAT ATGCTTGCCT TTGCACTGGG TACGCTGCCC AATCTTTTAG
 551 CAATCGGCAT TTTTCCCTG CAACTGAAA AAATCATGCA AAACCGATAT
 601 ATCCGCCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAACT
 651 TGCCGTCCTG TGGCTGTAA

This corresponds to the amino acid sequence <SEQ ID 394; ORF103-1>:

1 MNHDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPIH NREWLILLLN
 51 TGRVSSYTAI GLILGLIGQV GVSLDQTRVL QNILYTAANL LLLFLGLYLS
 101 GISSLAAKIE KIGKPIWRNL NPILNRLPI KSIPACLA VG ILWGWLP CGL
 151 VYSASLYALG SGSAATGGLY MLAFALGTL NLLAIGIFSL QLKIMQNR
 201 IRLCTGLSVS LWALWKLAVL WL*

Computer analysis of this amino acid sequence gave the following results:

15 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF103 shows 93.8% identity over a 222aa overlap with an ORF (ORF103a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
20	orf103.pep	MNHDITFLTLFLLGXF	GGTHCIGMCGGLSSAF	XXQLPPIHNR	EWLILLN	TGRVSSYTAI	
	orf103a	MNXDITFLTLFLLGFF	GGTHCIGMCGGLSSAF	ALQLPPIHNR	XWLILLN	TGRVSSYTAI	
		10	20	30	40	50	60
25	orf103.pep	GLILGLIGQVGVSLD	QTRVLQNILYTAANL	LLFLGLYLSGI	SSLAAKIEKIG	KPIWRNL	
	orf103a	GLILGLIGQVGVSLD	QTRVXQNILYTAANL	LLFLGLYLSGI	SSLAAKIEKIG	KPIWRNL	
		70	80	90	100	110	120
30	orf103.pep	NPILNRLPIKSIPAC	LA VGILWGWLP	CGLVYSASLYAL	SGSAATGGLY	MLAFALGTL	
	orf103a	NPILNRLPIKSIPAC	LA VGILWGWLP	CGLVYSASLYAL	SGSAATGGLY	MLAFALGTL	
		130	140	150	160	170	180
35	orf103.pep	NLLAIGIFSLQLX	KIMQNR	YIRLCTGLSV	SLWALWKLAVL	WLX	
	orf103a	NLXAIGIFSLQLX	KIMQNR	YIRLCTGLSV	SLWALWKLAVL	WLX	
		190	200	210	220		
40	orf103.pep	NLLAIGIFSLQLX	KIMQNR	YIRLCTGLSV	SLWALWKLAVL	WLX	
	orf103a	NLXAIGIFSLQLX	KIMQNR	YIRLCTGLSV	SLWALWKLAVL	WLX	
		190	200	210	220		

The complete length ORF103a nucleotide sequence <SEQ ID 395> is:

1 ATGAACCANG ACATCACTTT CCTCACCTG TTCCTACTCG GTTCTTTCGG
 51 CGGAACGCAC TGCATCGGTA TGTGCGCGG ATTAAGCAGC GCGTTTGC GC
 101 TCCAACCTCCC CCGCATATC AACCGCTTNT GGCTGATCCT GCTGCTTAAC
 45 151 ACAGGACGGG TAAGCAGCTA TACGGCAATC GGCCTGATAC TCGGATTAAT
 201 CGGACAGGTC GCGGTTTAC TCGACCAAC CCGCGTCNTG CAGAATATTT
 251 TATACACGGC CGCAACCTC CTGCTGCTCT TTTAGGCTT ATACTTGAGC
 301 GGTATTCTT CCTTGGCGGC AAAATCGAG AAAATCGCA AACCGATATG
 351 GCGGAACCTG AACCCGATAC TCAACCGGCT GTTACCCATA AAATCCATAC
 50 401 CCGCCTGCCT TCGCGTCGGA ATATTATGGG GCTGGCTGCC GTGCGGACTA
 451 GTTTACAGCG CGTCGCTTTA CGCGCTGGGA AGCGGTAGTG CGGCAACGGG
 501 CGGGTTATAT ATGCTTGCCT TTGCACTGGG TACGCTGCCC AATCTTTNGG
 551 CAATCGGCAT TTTTCCCTG CAACTGAAA AAATCATGCA AAACCGATAT
 601 ATCCGCCTGT GTACGGGATT ATCCGTATCA TTATGGGCAT TATGGAACT
 55 651 TGCCGTCCTG TGGCTGTAA

This encodes a protein having amino acid sequence <SEQ ID 396>:

1 MNXDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPIH NRXLILLLN
 51 TGRVSSYTAI GLILGLIGQV GVSLDQTRVX QNILYTAANL LLLFLGLYLS
 101 GISSLAAKIE KIGKPIWRNL NPILNRLPI KSIPACLA VG ILWGWLP CGL

151 VYSASLYALG SGSAATGGGLY MLAFALGTLP NLXAIGIFSL QLKKIMQNRV
201 IRLCTGLSVS LWALWKLAVL WL*

ORF103a and ORF103-1 show 97.7% identity in 222 aa overlap:

5	orf103a.pep	MNXDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRXWLILLNTGRVSSYTAI	10	20	30	40	50	60
	orf103-1	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFLILLNTGRVSSYTAI	10	20	30	40	50	60
10	orf103a.pep	GLILGLIGQVGVSLDQTRVXQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	70	80	90	100	110	120
	orf103-1	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	70	80	90	100	110	120
15	orf103a.pep	NPILNRLPIKSIPACLAVGILWGWLPCLVYSASLYALGSGSAATGGGLYMLAFALGTLP	130	140	150	160	170	180
	orf103-1	NPILNRLPIKSIPACLAVGILWGWLPCLVYSASLYALGSGSAATGGGLYMLAFALGTLP	130	140	150	160	170	180
20	orf103a.pep	NLXAIGIFSLQLKKIMQNRVIRLCTGLSVSLWALWKLAVLWLX	190	200	210	220		
	orf103-1	NLLAIGIFSLQLKKIMQNRVIRLCTGLSVSLWALWKLAVLWLX	190	200	210	220		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF103 shows 95.5% identity over a 222aa overlap with a predicted ORF (ORF103.ng) from *N.*

30 *gonorrhoeae*:

30	orf103.pep	MNHDITFLTLFLLGXFGGTHCIGMCGGLSSAFXXQLPPHINRFLILLNTGRVSSYTAI	60
	orf103ng	MNHDITFLTLFLLGFFGGTHCIGMCGGLSSAFALQLPPHINRFLILLNTGRVSSYTAI	60
35	orf103.pep	GLILGLIGQVGVSLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	120
	orf103ng	GLMLGLIGQLGISLDQTRVLQNILYTAANLLLLFLGLYLSGISSLAAKIEKIGKPIWRNL	120
40	orf103.pep	NPILNRLPIKSIPACLAVGILWGWLPCLVYSASLYALGSGSAATGGGLYMLAFALGTLP	180
	orf103ng	NPILNRLPIKSIPACLAVGILWGWLPCLVYSASLYALGSGSATTGGGLYMLAFALGTLP	180
45	orf103.pep	NLLAIGIFSLQLKKIMQNRVIRLCTGLSVSLWALWKLAVLWL	222
	orf103ng	NLLAIGIFSLQLKKIMQNRVIRLCTGLSVSLWALWKLAVLWL	222

The complete length ORF103ng nucleotide sequence <SEQ ID 397> is:

1	ATGAACCACG	ACATCACTTT	CCTCACCCCTG	TTCCTGCTCG	GTTTCTTCGG
51	CGGAACAC	TGCATCGGTA	TGTGCGGCGG	ATTAAGCAGC	GCGTTTGCGC
101	TCCAACCTCC	CCCGCATATC	AACCGCTTTT	GGCTGATTCT	GCTGCTTAAC
151	ACAGGACGGA	TAAGCAGCTA	TACGGCAATC	GGCCTGATGC	TCGGATTAAT
201	CGGACAACCTC	GGCATTTCAC	TCGACCAAC	ccgcgTCCTG	CAAAATATTT
251	tatacacagc	ctccaaCCTC	CTGCTGCTCT	TTTTAGGCTT	ATACTTGAGC
301	GGTATTTCTT	CCTTGGCGGC	AAAAATCGAG	AAAATCGGCA	AACCGATATG
351	GCGCAACCTG	AACCCGATAC	TCAACCGGCT	GCTGCCCCATA	AAATCCATAC
401	CCGCCTGCCT	TGCTGTCCGA	ATATTATGGG	GCTGGCTGCC	GTGCGGACTG
451	GTTTACAGCG	CATCACTTTA	CGCGCTGGGA	AGCGGTAGTG	CGACAACCGG
501	CGGACTGTAT	ATGCTTGCCT	TTGCACTGGG	TACGCTGCCC	AATCTTTTGG
551	CAATCGGCAT	TTTTTCCTTG	CAACTGAAAA	AAATCATGCA	AAACCGATAT
601	ATCCGCCTGT	GTACAGGATT	ATCCGTATCA	TTATGGGCAT	TATGGAAGCT
651	TGCCGTCCTG	TGGCTGTAA			

This encodes a protein having amino acid sequence <SEQ ID 398>:

1 MNHDITFLTL FLLGFFGGTH CIGMCGGLSS AFALQLPPHI NRFWLILLN
 51 TGRISSYTAI GLMLGLIGQL GISLDQTRVL QNILYTASNL LLLFLGLYLS
 101 GISSLAAKIE KIGKPIWRNL NPILNRLPI KSIPACLA VLGILWGLPCGL
 151 VYSASLYALG SGSATTGGLY MLAFALGTL PNLLAIGIFSL QLKKIMQNR
 201 IRLCTGLSVS LWALWKLAVL WL*

In addition, ORF103ng and ORF103-1 show 97.3% identity in 222 aa overlap:

		10	20	30	40	50	60
orf103-1.pep		MNHDITFLTL	FLLGFFGGTH	CIGMCGGLSS	AFALQLPPHI	NRFWLILLN	TGRVSSYTAI
orf103ng		MNHDITFLTL	FLLGFFGGTH	CIGMCGGLSS	AFALQLPPHI	NRFWLILLN	TGRVSSYTAI
		10	20	30	40	50	60
		70	80	90	100	110	120
orf103-1.pep		GLILGLIGQV	GVSLDQTRVL	QNILYTAN	LLLLFLGLY	LSGISSLA	AKIEKIGKPIWRNL
orf103ng		GLMLGLIGQL	GISLDQTRVL	QNILYTAN	LLLLFLGLY	LSGISSLA	AKIEKIGKPIWRNL
		70	80	90	100	110	120
		130	140	150	160	170	180
orf103-1.pep		NPILNRLPI	KSIPACLA	VGILWGLP	CGLVYSAS	LYALGSGS	AATGGLYMLAFALGTL
orf103ng		NPILNRLPI	KSIPACLA	VGILWGLP	CGLVYSAS	LYALGSGS	AATGGLYMLAFALGTL
		130	140	150	160	170	180
		190	200	210	220		
orf103-1.pep		NLLAIGIFSL	QLKKIMQNR	YIRLCTGL	SVSLWALW	KLAVLWLX	
orf103ng		NLLAIGIFSL	QLKKIMQNR	YIRLCTGL	SVSLWALW	KLAVLWLX	
		190	200	210	220		

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 47

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 399>:

1 ATGGAAAACC AAAGGCCGCT C TAGGCTTT CGTTGGCAC TTTGGCGGC
 51 GATGACGTGG GGAACGCTGC CGAT.TCCGT GCGGCAGGTA TTGAAGTTTG
 101 TCGATGCGCC GACGCTGGTG TGGGTGCGTT TTACCGTGGC GCGCGCGGTA
 151 TTGTTTGTTT TGCTGGCACT GGGCGGGCGG CTGCGGAAGC GCGCGGATT
 201 TTTCTTGCTG CTCATTGAGG CTGCTGCTGC TCGGCGTGGC GGGCATTTCG
 251 GCAAACTTTG TGCTGATTGC CCAAGGGCTG CATTATATTT CGCCGACCAC
 301 GACGCAGGTT TTGTGGCAGA TTTCGCCGTT TACGATGATT GTWGTGCGGT
 351 TGTTGGTGTT TAAAGACCGG ATGACTGCCG CTCAGAAAAT CGGCTTGCTT
 401 TTGCTGCTTG CCGGTTTGCT TATGTATTT AACGATAAAT TCGGCGAGTT
 451 GTCGGGTTTG GCGCGGTATG C.AAGGGCGT GTTGCTGTGT GCGGCAGGCA
 501 GTATGGCATG GGTGTGTAAT GCCGTGGCGC AAAAGCTGCT GTCGGCGCAA
 551 TTCGGGCGCG AACAGATTCT GCTGTTGATT TATGCGGCAA GTCCCGCGCT
 601 GTTCCTGCGG TTTGCCGAAC CGGCACACAT CGGAAGTATG GACGGTACGT
 651 TGGCGTGGGT ATGTATTGCG TATTGCTGCT TGAATACGTT AATCGGTTAC
 701 GGCTCGTTTC GCGAGGCGTT GAAACATTGG GAGGCTTCCA AAGTCAGCGC
 751 GGTAACAACC TTGCTCCCGG TGTTACCGT AATAAATACT TTGCTCGGGC
 801 ATTATGTGAT GCCTGAAACT TTTGCCGCGC CGGA..

This corresponds to the amino acid sequence <SEQ ID 400; ORF104>:

1 MENQRPLLGF RLALLAAMTW GTLPXSVRQV LKFVDAPTLV WVRFTVAAAV
 51 LFLVLLALGGR LPKRRDFSWC SFRLLLLGVA GISANFVLIA OGLHYISPTT
 101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLAGLL MYFNDKFGEI
 151 SGLGAYXKGV LLCAAGSMAW VCNAAQKLL SAQFGPQOIL LLIYAASAAV
 201 FLPFAEPAHI GSMDGTLAWV CIAYCLNTL IGYGSFGEAL KHWEASKVSA

Further work revealed further partial DNA sequence <SEQ ID 401>:

	1	ATGGA ⁺ AAACC	AAAGGCGCCT	CCTAGGCTTC	GCCTTGGCAC	TTTTGGCGGC
5	51	GATGACGTGG	GGAACGCTGC	CGATTGCCGT	GCGGCAGGTA	TTGAAGTTTG
	101	TCGATCGCGC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGGC	GGCGCGGTA
	151	TTGTTTGT ⁺ TT	TGCTGGCATT	GGCGGGCGCG	CTGCCGAAGC	GGCGGGATT ⁺ T
	201	TTCTTGGTGC	TCATTACAGC	TGCTGCTGCT	CGGCGTGGCG	GGCATTTCGG
10	251	CAAACTTTGT	GCTGATTGCC	CAAGGGCTGC	ATTATATTTC	GCCGACCACG
	301	ACGCAGGTTT	TGTGGCAGAT	TTCCGCGTTT	ACGATGATTG	TTGTCCGTTG
	351	GTTGGTGT ⁺ TT	AAAGACCGGA	TGACTGCCGC	TCAGAAAATC	GGCTTGTT ⁺ TT
	401	TGCTGCTTGC	CGGTTTGCTT	ATGTTTTTTA	ACGATAAATT	CGGCGAGTTT
15	451	TCGGGTTTGG	GC ⁺ CGCTATGC	GAAGGCGGTG	TTGCTGTGTG	CGGCAGCGAG
	501	TATGGCATGG	GTGTGTTATG	CCGTGGCGCA	AAAGCTGCTG	TCGGCGCAAT
	551	TCGGGCGGCA	ACAGATTCTG	CTGTTGATTT	ATCGCGCAAG	TGCCCGCGTG
	601	TTCTTGCCGT	TTGCCGAACC	GGCACACATT	GGAAGTTTGG	ACGGTACGTT
	651	GGCGTGGGTT	TGTTT ⁺ TGCGT	ATTGCTGCTT	GAATACGTTA	ATCGGTTACG
	701	GCTCGTTCCG	CGAGGCGTTG	AAACATTGGG	AGGCTTCCAA	AGTCAGCGCG
	751	GTAACAACCT	TGCTCCCCGT	GTTTACCCTA	ATAwTwwCTT	TGCTCGGGCA
	801	TTATGTGATG	CTGAAACTT	TTGCCGCGCC	AGA...	

20 This corresponds to the amino acid sequence <SEQ ID 402; ORF104-1>:

25

1	<u>MENQRPLLGF</u>	<u>ALALLAAMTW</u>	<u>GTLPIAVRQV</u>	<u>LKFVDAPTLV</u>	<u>WVRFTVAAAV</u>
51	<u>LFVLWLLGGR</u>	<u>LPKRRDFSWC</u>	<u>SFRLLLLGVA</u>	<u>GISANFVLIA</u>	<u>QGLHYISPTT</u>
101	<u>TOVLWQISPF</u>	<u>TMIVVGVLVF</u>	<u>KDRMTAAQKI</u>	<u>GLVLLLAGLL</u>	<u>MFNDKFGTEL</u>
151	<u>SGLGAYAKGV</u>	<u>LLCAAGSMAW</u>	<u>VCYAVAQKLL</u>	<u>SAQFGPQQIL</u>	<u>LLIYAASAAV</u>
201	<u>FLPFAEPAHI</u>	<u>GSLDGTLAW</u>	<u>CFAYCCLNTL</u>	<u>IGYGSFGEAL</u>	<u>KHWEASKVSA</u>
251	<u>VTLLPVFTV</u>	<u>IXXLLGHYVM</u>	<u>PETFAAP...</u>		

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical HI0878 protein of *H. influenzae* (accession number U32769)

ORF104 and HI0878 show 40% aa identity in 277aa overlap:

30	orf104	4	QRPLLGFRLALLAAMTWGTLPKXSVRQVLKFVDAPTLVWXXXXXXXXXXXXXXXXXXXXXP- 62
			Q+PLLG F AL+ AM WG+LP +++QVL ++A T+VW P
	HI0878	3	QQPLLGFTEALITAMAWGSLPIALKQVLSVMNAQITVWYRFIIAAVSLALLAYKKQLPE 62
35	orf104	63	--KRRDFSWCFSRLLLLGVAGISANFVLIAOGLHYISPTTTQVLWQISPFMTIVVGVLVF 120
			K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+F
	HI0878	63	LMKVRQYAW----IMLIGVIGLTSNFLLFSSSLNYIEPSVAQIFIHLSSFGMLICGVLI F 118
40	orf104	121	KDRMTAAQKIXXXXXXXXXXXMYFNDKFGELSGLGAYXKGVLCCAAGSMAWVCNAVAQKLL 180
			K+++ QKI ++FND+F +GL Y GV+L G++ WV +AQKL+
	HI0878	119	KEKLGLHQKIGLFLLLIGLGLFNFDRDFAAGLNGQYSTGVILGVGGALIWVAYGMAQKLM 178
45	orf104	181	SAQFGPQQIILLIYAASAAVFLPFAEPAHIGSMDGTLAWVCIA YCCLNTLIGYGSFGEAL 240
			+F QQILL++Y A F+P A+ + + + LA +C YCCLNTLIGYGS+ EAL
	HI0878	179	LRKFNSQQIILLMYLGCIAFMFMADFSQVQELT-PLALICFIYCCLNTLIGYGSYAEAL 237
	orf104	241	KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP 277
			W+ SKVS V TL+P+FT++ + + HY P FAAP
	HI0878	238	NRWDVSKSVVITLVPLFTILFSHIAHYFSPADFAAP 274

50 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF104 shows 95.3% identity over a 277aa overlap with an ORF (ORF104a) from strain A of *N.*

meningitidis:

55

orf104.pep	MENQRPLLGFRLLAALLAAMTWGTL	PXSRQVLK	FVDAPTLVWVRFTVAAAVLFVLLALGGR
orf104a	MENQRPLLGFRLLAALLAAMTWGTL	PIAVRQVLK	FVDAPTLVWVRFTVAAAVLFVLLALGGR

-251-

5	orf104.pep	LPKRRDFSWCSFRLLLLGVAGISANFVLIAOGLHYISPTTQVLWQISPFMTMIVVGVLF
	orf104a	LPKWRDFSWCSFRLLLLGVAGISANFVLIAOGLHYISPTTQVLWQISPFMTMIVVGVLF
10	orf104.pep	KDRMTAAQKIGLVLLLAGLLMYFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL
	orf104a	KDRMTAAQKIGLVLLLAGLLMEFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL
15	orf104.pep	SAQFGPQOILLIYAASAAVFLPFAEPAHIGSMDGT LAWVCIAYCCLNTLIGYGSFGEAL
	orf104a	SAQFGPQOILLIYAASAAVFLPFAELAHIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL
20	orf104.pep	KHWEASKVSAVTLLPVFTVINTLLGHYVMPETFAAP
	orf104a	KHWEASKVSAVTLLPVFTVIFSLGHYVMPDTFAAPDMNGLGYAGALVVVGAVTAAVG

The complete length ORF104a nucleotide sequence <SEQ ID 403> is:

1	ATGGAACC	AAAGCCGCT	CCTAGGCTTC	CGCTGGCAC	TTTTGGCGGC
51	GATGACGTGG	GGAACGCTGC	CGATTGCCGT	GCGGCAGSTA	TTGAAGTTTG
101	TCGATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGGC	GGCGGCGGTA
151	TTGTTTGT	TGCTGGCATT	GGCGGGCGG	CTGCCGAAGT	GGCGGGATT
201	TTCTTGCTGC	TCATTGAGGC	TGCTGCTGCT	CGCGTGCGG	GGCATTTCGG
251	CAAACTTTGT	GCTGATTGCC	CAAGGCTGC	ATTATATTTC	GCCGACCACG
301	ACGCAGGTTT	TGTGGCAGAT	TCGCCGTTT	ACGATGATTG	TTGTCGGTGT
351	GTTGGTGTTT	AAAGACCGGA	TGACTGCCGC	TCAGAAAATC	GGCTTGTTT
401	TGCTGCTTGC	CGGTTTGCTT	ATGTTTMTTA	ACGATAAATT	CGGCAGTTG
451	TCGGGTTTGG	GCGCGTATGC	GAAGGCGTG	TTGCTGTGTG	CGGCAGGCAG
501	TATGGCATGG	GTGTGTATG	CCGTGGCGCA	AAAGCTGCTG	TCGGCGCAAT
551	TCGGGCGGCA	ACAGATTCTG	CTGTTGATT	ATGCGGCAAG	TGCCGCGTG
601	TTCTTGCCGT	TGCGGAACT	GGCACACATC	GGAAGTTTGG	ACGGTACGTT
651	GGCGTGGGTT	TGTTTTCGCT	ATTGCTGCTT	GAATACGTTA	ATCGGTTACG
701	GCTCGTTCGG	CGAGGCGTTG	AAACATTGGG	AGGCTTCCAA	AGTCAGCGCG
751	GTAACAACCT	TGCTCCCGGT	GTTACCGTA	ATATTTCCTT	TGCTCGGGCA
801	TTATGTGATG	CCTGATACTT	TTGCCGCGCC	GGATATGAAC	GGTTTGGGTT
851	ATGCCGGCGC	ACTGGTCGTG	GTCGGGGGTG	CGGTTACGGC	GGCGGTGGGG
901	GACAGGCTGT	TCAAACGCCG	CTAG		

This encodes a protein having amino acid sequence <SEQ ID 404>:

1	MENQRPLLGF	ALALLAAMTW	GTLPVAVRQV	LKFVDAPTLV	WVRFTVAAAV
51	LFVLLALGGR	LPKWRDFSWC	SFRLLLLGVA	GISANFVLIA	QGLHYISPTT
101	TQVLWQISPF	TMIVVGVLF	KDRMTAAQKI	GLVLLLAGLL	MEFNDKFGEL
151	SLGAYAKGV	LLCAAGSMAW	VCYAVAQKLL	SAQFGPQOIL	LLIYAASAAV
201	FLPFAELAH	ISLDGT LAWV	CFAYCCLNTL	IGYGSFGEAL	KHWEASKVSA
251	VTTLLPVFTV	IFSLGHYVM	PDTFAAPDMN	GLGYAGALVV	VGGAVTAAVG
301	DRLEKRR*				

ORF104a and ORF104-1 show 98.2% identity in 277 aa overlap:

55	orf104a.pep	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR
	orf104-1	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR
60	orf104a.pep	LPKWRDFSWCSFRLLLLGVAGISANFVLIAOGLHYISPTTQVLWQISPFMTMIVVGVLF
	orf104-1	LPKRRDFSWCSFRLLLLGVAGISANFVLIAOGLHYISPTTQVLWQISPFMTMIVVGVLF
65	orf104a.pep	KDRMTAAQKIGLVLLLAGLLMEFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL

-252-

	orfl04-1	 KDRMTAAQKIGLVLLLAGLLMFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL 130 140 150 160 170 180
5	orfl04a.pep	190 200 210 220 230 240 SAQFGPQQIILLIYAASAAVFLPFAELAHIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL
10	orfl04-1	190 200 210 220 230 240 SAQFGPQQIILLIYAASAAVFLPFAEPAHIGSLDGT LAWVCFAYCCLNTLIGYGSFGEAL
	orfl04a.pep	250 260 270 280 290 300 KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLGYAGALVVVGAVTAAVG
15	orfl04-1	250 260 270 KHWEASKVSAVTTLLPVFTVIXLLGHYVMPETFAAP

Homology with a predicted ORF from *N.gonorrhoeae*

ORF104 shows 93.9% identity over a 277aa overlap with a predicted ORF (ORF104.ng) from *N. gonorrhoeae*:

20	orfl04.pep	MENQRPLLGFRLALLAAMTWGTLPXSVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
	orfl04ng	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFVLLALGGR	60
25	orfl04.pep	LPKRDFSWSFRLLLLGVAGISANFVLIAQGLHYISPSTTTQVLWQISPFMTMIVGVLVF	120
	orfl04ng	LPKRDFSWSFRLLLLGVAGISANFVLIAQGLHYISPSTTTQVLWQISPFMTMIVGVLVF	120
30	orfl04.pep	KDRMTAAQKIGLVLLLAGLLMFNDKFGELSGLGAYXKGVLLCAAGSMAWVCNAVAQKLL	180
	orfl04ng	KDRMTAAQKIGLVLLLAGLLMFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL	180
35	orfl04.pep	SAQFGPQQIILLIYAASAAVFLPFAEPAHIGSMDGT LAWVCIAYCCLNTLIGYGSFGEAL	240
	orfl04ng	SAQFGPQQIILLIYAASAAVFLXAEPAHIGSLDGT LAWVCFVYCCLNTLIGYGSFGEAL	240
	orfl04.pep	KHWEASKVSAVTTLLPVFTVINTLLGHYVMPETFAAP	277
	orfl04ng	KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLGYVVALVVVGAVTAAVG	300

The complete length ORF104ng nucleotide sequence <SEQ ID 405> is predicted to encode a protein having amino acid sequence <SEQ ID 406>:

1	MENQRPLLGF ALALLAAMTW	GTLPIAVRQV	LKFVDAPTLV	WVRFTVAAAV
51	LFVLLALGGR LPKRDFSWSH	SFRLLLLGVT	GISANFVLIA	QGLHYISPST
101	TQVLWQISPF TMIVGVVLVF	KDRMTAAQKI	GLVLLLVGLL	MFNDKFGEL
151	SGLGAYAKGV LLCAAGSMAW	VCYAVAQKLL	SAQFGPQQIL	LLIYAASAAV
201	FLLXAEPAHI GSLDGT LAWV	CFVYCCLNTL	IGYGSFGEAL	KHWEASKVSA
251	VTLLPVFTV IFSLLGHYVM	PDTFAAPDMN	GLGYVVALVV	VGGAVTAAVG
301	DRPFKRR*			

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 407>:

50	1	ATGGA AAC	AAAGGCCGCT	CCTAGGCTTC	GCGTTGGCAC	TTTGGCGGC
	51	GATGACGTGG	GGGACGCTGC	CGATTGCCGT	GCGGCAGGTA	TTGAAGTTTG
	101	TCGATGCGCC	GACGCTGGTG	TGGGTGCGTT	TTACCGTGGC	GGCGGCGGTA
	151	TTGTTTGT	TGCTGGCATT	GGGCGGGCGG	CTGCCGAAGC	GGCGGGATT
	201	TTCTGGCAT	TCATT CAGC	TGCTGCTGCT	CGGCGTGACG	GGCATTTCGG
	251	CAAACCTTGT	GCTGATTGCC	CAAGGGCTGC	ATTATATTTC	GCCGACCACG
55	301	ACGCAGGTTT	TGTGGCAGAT	TTGCGCGTTT	ACGATGATTG	TTGTCGGCGT
	351	GTTGGTGT	AAAGACCGGA	tgaCTGCCG	GCAGAAAATC	GTTTGGTTT
	401	TGCTGCTgt	CGTttgCTT	ATGTTTTta	ACGACAAAT	CGGCGAGTTG
	451	TCGGTTTGG	GCGCGTATGC	GAAGGGCGTG	TTGCTGTGTG	CGGCAGGCAG
	501	TATGGCCTGG	GTGTGTTATG	CCGTGGCGCA	AAAGCTGCTG	TCCGCGCAAT
60	551	TCGGGCGGCA	ACAGATTCTG	CTGTTGATT	ATGCGGcaag	tgccgccGTG
	601	TTCctgccgt	TTGccgaaCC	GGCACACATC	GGAAGTTTgg	aCGGTACGtt
	651	GGCGTGGGTT	TGTTTTGTGT	ATTGCTGCTT	GAATACGTTA	ATCGGTTACG

-253-

701 GCTCGTTCGG CGAGGCGTTG AACATTGGG AGGCTTCCAA AGTCAGCGCG
 751 GTAACAACCT TGCTCCCCGT GTTACCGTA ATATTTTCTT TGCTCGGGCA
 801 TTATGTGATG CCTGATACTT TTGCCGCGCC GGATATGAAC GGTTTGGGT
 851 ATGTCGGCGC ACTGGTCGTG GTCGGGGGTG CCGTTACGGC GCGGGTGGGG
 901 GACAGGCCGT TCAAACGCCG CTAG

This corresponds to the amino acid sequence <SEQ ID 408; ORF104ng-1>:

1 MENQRPLLGF ALALLAAMTW GTLPIAVRQV LKFVDAPTLV WVRFTVAAAV
 51 LFLVLLALGGR LPKRRDFSWH SFRLLLLGVT GISANFVLIA QGLHYISPTT
 101 TQVLWQISPF TMIVVGVLVF KDRMTAAQKI GLVLLLVGLL MFFNDKFGEL
 151 SGLGAYAKGV LLCAAGSMAW VCYAVAQKLL SAQFGPQQIL LLIYAASAAV
 201 FLPPFAEPAHI GSLDGT LAWV CFVYCCLNTL IGYGSFGEAL KHWEASKVSA
 251 VTLLPVFTV IFSLLGHYVM PDTFAAPDMN GLGYVGALVV VGGAVTAAVG
 301 DRPFKRR*

ORF104ng-1 and ORF104-1 show 97.5% identity in 277 aa overlap:

15	orf104-1.pep	10	20	30	40	50	60
	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFLVLLALGGR						
	orf104ng-1	10	20	30	40	50	60
	MENQRPLLGFALALLAAMTWGTLPIAVRQVLKFVDAPTLVWVRFTVAAAVLFLVLLALGGR						
20	orf104-1.pep	70	80	90	100	110	120
	LPKRRDFSWCSFRLLLLGVAGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF						
	orf104ng-1	70	80	90	100	110	120
	LPKRRDFSWHSFRLLLLGVGTGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF						
25	orf104-1.pep	130	140	150	160	170	180
	KDRMTAAQKIGLVLLLAGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL						
	orf104ng-1	130	140	150	160	170	180
	KDRMTAAQKIGLVLLLVGLLMFFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL						
30	orf104-1.pep	190	200	210	220	230	240
	SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGT LAWVCFYCCCLNTLIGYGSFGEAL						
	orf104ng-1	190	200	210	220	230	240
	SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGT LAWVCFYCCCLNTLIGYGSFGEAL						
35	orf104-1.pep	250	260	270			
	KHWEASKVSAVTTLLPVFTVIXXLLGHYVMPETFAAP						
	orf104ng-1	250	260	270	280	290	300
	KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMNGLGYVGALVVVGGAVTAAVG						

In addition, ORF104ng-1 shows significant homology with a hypothetical *H. influenzae* protein:

45 gi|1573895 (U32769) hypothetical [Haemophilus influenzae] Length = 306
 Score = 237 bits (598), Expect = 8e-62
 Identities = 114/280 (40%), Positives = 168/280 (59%), Gaps = 8/280 (2%)

50 Query: 30 QRPXXXXXXXXXXMTWGTLPPIAVRQVLKFVDAPTLVWXXXXXXXXXXXXXXXXXXXXXP- 88
 Q+P M WG+LPIA++QVL ++A T+VW P
 Sbjct: 3 QPPLLGFALITAMAGSLPIALKQVLSVMNAQTIVWYRFIIAAVSLALLAYKKQLPE 62

55 Query: 89 --KRRDFSWHSFRLLLLGVGTGISANFVLIAQGLHYISPTTTQVLWQISPFTMIVVGVLVF 146
 K R ++W ++L+GV G+++NF+L + L+YI P+ Q+ +S F M++ GVL+
 Sbjct: 63 LMKVRQYAW----IMLIGVIGLTSNFLFSSSLNIEPSVAQIFIHLSSFGMLICGVLF 118

60 Query: 147 KDRMTAAQKIXXXXXXXXXXMFNDKFGELSGLGAYAKGVLLCAAGSMAWVCYAVAQKLL 206
 K+++ QKI +FFND+F +GL Y+ GV+L G++ WV Y +AQKL+
 Sbjct: 119 KEKLGHLQKIGLFLLLIGLGLFFNDRFADFAGLNQYSTGVILGVGGALIWVAYGMAQKLM 178

65 Query: 207 SAQFGPQQILLIYAASAAVFLPFAEPAHIGSLDGT LAWVCFYCCCLNTLIGYGSFGEAL 266
 +F QQILL++Y A F+P A+ + + L LA +CF+YCCCLNTLIGYGS+ EAL
 Sbjct: 179 LRKFNSQQILLMMYLGCIAIFMPMAFDSQVQELT-PLALICFTYCCCLNTLIGYGSYAEAL 237

Query: 267 KHWEASKVSAVTTLLPVFTVIFSLLGHYVMPDTFAAPDMN 306

W+ SKVS V TL+P+FT++FS + HY P FAAP++N
 Sbjct: 238 NRWDVSKVSVVITLVPLFTILFSHIAHYFSPADFAAPELN 277

Based on this analysis, including the presence of a putative leader sequence and several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from
 5 *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 48

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 409>:

```

10      1 ATGGTAGCTC GTCGGGCTCA TAACCCGAAG GTCGTAGGTT CGAATCCTGT
      51 .CCCGCAACC TAATTTCAAA CCCCTCGGTT CAATGCCGAG GG.GTTTTGT
    101 T.TTGCCTGT TTCCTGTTTC CTGTTTCCTG CCGCCTCCGT TTTTGGCCGG
    151 ATTTTCCTTC CGGCCGCAAT ATCGGAACGG CAGACCGCCG TCTGTTTGGC
    201 GTTGCAAATT CAGGCAGTTT GGCTACAATC TTCCGCATTG TCTTCAAGAA
    251 AGCCAAACCAT GCCGACCGTC CGTTTACCG AATCCGTCAG CAAACAAGAC
    15 301 CTTGATGCTC TGTTTCAGTG GGCAAAAGCA AGTTACGGTG CAGAAAGTTG
    351 CTGGAAAACG CTGTATCTGA ACGGTCysCC TTTGGGCAAC CTGTCGCCGG
    401 AATGGGTGGA ACGCGTsmmA AAAGACTGGG AGGCAGGCTG CyCGGAGTCT
    451 TCAGACGGCA TTTTCTGAA TgCGGACGGc TgGcctGATA TGGgCGGAcg
    501 cTTACAGCAC CTCGCCCTCG GTTGGCACTG TCGGGGGCTG TTGGACGgsT
    20 551 GGCGCAACGA GTGTTTCGAC CTGACCGACG GCGGCGGCAA CCCCTTGTTT
    601 ACGCTCGaAc GCGCCGyTTT mCGTCCTkTC GGA CTGCTCA GCCGCGCCGT
    651 CCATCTCAAC GGTCTGACCG AATCGGACGG CCGATGGCAT TTCTGGATAG
    701 GCAGGCGCAG TCCGCACAAA GCAGTCGATC CCAACAAACT CGACAATACT
    751 rCCGCCGGCG GTGTTTCCGG CGGCGAAATG CCGTCTGAAG CCGTGTGTCTG
    25 801 CGAAAGCAGC GAAGAAGCCG GTTTGGATAA AACCTGcTT CCGCTCATCC
    851 GCCCGGTATC GCAGCTGCAC AGCCTGCGCT CCGTCAGCCG GGGTGTACAC
    901 AATGAAATCC TGTATGTATT CGATGCCGTC CTGCCG...
```

This corresponds to the amino acid sequence <SEQ ID 410; ORF105>:

```

30      1 MVARRAHNPV VVGSNPXPAT XFQTPRFNAE XVLXLPVSCF LFPAASVFRCR
      51 IFLPAAISER QTAVCLRLQI QAVWLQSSAL SSRKPTMPTV RFTESVSKQD
    101 LDALFEWAKA SYGAESCWKT LYLNGXPLGN LSPWVERVX KDWEAGCXES
    151 SDGIFLNADG WPDMMGRLQH LALGWHCAGL LDGWRNECFD LTDGGGNPLF
    201 TLERAXRXPX GLLSRAVHLN GLTESDGRWH FWIGRRSPHK AVDPNKLNT
    251 XAGGVSGGEM PSEAVCRESS EEAGLDKTLT PLIRPVSQLH SLRSVSRGVH
    35 301 NEILYVFDV LP...
```

Further work revealed the complete nucleotide sequence <SEQ ID 411>:

```

40      1 ATGCCGACCG TCCGTTTTAC CGAATCCGTC AGCAAACAAG ACCTTGATGC
      51 TCTGTTTCGAG TGGGCAAAAG CAAGTTACGG TGCAGAAAGT TGCTGGAAAA
    101 CGCTGTATCT GAACGGTCTG CCTTTGGGCA ACCTGTCGCC GGAATGGGTG
    151 GAACGCGTCA AAAAAGACTG GGAGGCAGGC TGCTCGGAGT CTTCAGACGG
    201 CATTTTTCTG AATGCGGACG GCTGGCCTGA TATGGGCGGA CGCTTACAGC
    251 ACCTCGCCCT CGGTTGGCAC TGTGCGGGC TGTGGACGG CTGGCGCAAC
    301 GAGTGTTCG ACCTGACCGA CGGCGGCGGC AACCCTTGT TCACGCTCGA
    351 ACGCGCCGCT TTCGTCCTT TCGGACTGCT CAGCCGCGCC GTCCATCTCA
    45 401 ACGTCTGAC CGAATCGGAC GGCCGATGGC ATTTCTGGAT AGGCAGGCGC
    451 AGTCCGCACA AAGCAGTCGA TCCCAACAAA CTCGACAATA CTGCCGCCGG
    501 CGGTGTTTCC GGCGGCGAAA TGCCGTCTGA AGCCGTGTGT CGCGAAAGCA
    551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGC TTCCGCTCAT CCGCCCGGTA
    601 TCGCAGCTGC ACAGCCTGCG CTCCGTCAGC CGGGGTGATC ACAATGAAAT
    50 651 CCTGTATGTA TTCGATGCCG TCCTGCCCGA AACCTTCCTG CCGTGAATC
    701 AGGATGGCGA AGTGGCGGGT TTTGAGAAAA TGGACATCGG CCGTCTGTTG
    751 GATGCCATGT TGTCGGGAAA CATGATGCAC GACGCGCAAC TGGTTACGCT
    801 GGACGCGTTT TGCCGTTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG
    851 AGTGGCTGGA CGGCATACGT TTATAG
```

55 This corresponds to the amino acid sequence <SEQ ID 412; ORF105-1>:

-255-

1 MPTVRFTESV SKQDLDALE WAKASYGAES CWKTLYLNGL PLGNLSPEWV
 51 ERVKKDWEAG CSESSDGIFL NADGWPMGG RLQHLALGWH CAGLLDGWRN
 101 ECFDLTDGGG NPLFTLERAA FRPFGLLSRA VHLNGLTESD GRWHFWIGRR
 151 SPHKAVDPNK LDNTAAGGVS GGEMPSEAVC RESSEEAGLD KTLPLIRPV
 201 SQLHSLRSVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGGLL
 251 DAMLSGNMMH DAQLVTLDAF CRYGLIDAHA PLSEWLDGIR L*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF105 shows 89.4% identity over a 226aa overlap with an ORF (ORF105a) from strain A of *N.*

10 *meningitidis*:

	60	70	80	90	100	110
orf105.pep	ISERQTAVCLRLQIQAVWLQSSALSSRKPTMPTVRFTESVSKQDLDALEWAKASYGAES					
orf105a	MPTVRFTESVSKHDLDALEWAKASYGAES					
				10	20	30
orf105.pep	120	130	140	150	160	170
	CWKTLYLNGXPLGNLSPEWVERVKKDWEAGCXESSDGIFLNADGWPMGGRLQHLALGWH					
orf105a	CWKTLYLNGLPLGNLSPEWAERVKKDWEAGCXESSDGIFLNADGWPMGGRLQHLARIWK					
	40	50	60	70	80	90
orf105.pep	180	190	200	210	220	230
	CAGLLDGWRNECFDLTDGGGNPLFTLERAXRPXGLLSRAVHLNGLTESDGRWHFWIGRR					
orf105a	EAGLLHGWRDECFDLTDGGSNPLFALERAARFPFGLLSRAVHLNGLVESDGRWHFWIGRR					
	100	110	120	130	140	150
orf105.pep	240	250	260	270	280	290
	SPHKAVDPNKLDNTXAGGVSGGEMPSEAVCRESSEEAGLDKTLPLIRPVSQHLHSLRSVS					
orf105a	SPHKAVDPDKLDNTAAGGVSSGELPSETVCRESSEEAGLDKTLPLIRPVSQHLHSLRPVS					
	160	170	180	190	200	210
orf105.pep	300	310				
	RGVHNEILYVFDVLP					
orf105a	RGVHNEILYVFDVLPETFLPENQDGEVAGFEKMDIGGLLAAMLSGNMMHDAQLVTLDAF					
	220	230	240	250	260	270

40 The complete length ORF105a nucleotide sequence <SEQ ID 413> is:

1 ATGCCGACCG TCCGTTTAC CGAATCCGTC AGCAAACACG ACCTTGATGC
 51 CCTATTGAG TGGGCAAAGG CAAGTTACGG TGCGGAAAGT TGCTGAAAA
 101 CGCTGTATCT GAACGGTCTG CCTTTGGGCA ATCTGTCGCC GGAATGGGCG
 151 GAGCGCGTCA AAAAAGACTG GGAGGCAGGC TGCTCGGAGT CTTGACACGG
 201 CATTTCCTG AATGCGGACG GCTGGCCAGA TATGGGCAGA CGCTGCAGC
 251 ACCTCGCCCG AATATGGAAG GAAGCGGGAC TGCTTCACGG CTGGCGCGAC
 301 GAGTGTTTCG ACCTGACCGA CGGCGGCAGC AATCCCTTGT TCGCGCTCGA
 351 ACGCGCCGCT TTCCGTCCGT TCGGACTGCT CAGCCGCGCC GTCCATCTCA
 401 ACGGTTTGGT CGAATCGGAC GGCCGATGGC ATTTCTGGAT AGGCAAGCGC
 451 AGTCCGCACA AAGCAGTCGA TCCCGACAAA CTCGACAATA CTGGCGCCGG
 501 CGGTGTTTCC AGCGGTGAAT TGCCGTCTGA AACCGTGTGT CGCGAAAGCA
 551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGC TTCCGCTCAT CCGCCGGTA
 601 TCGCAGCTGC ACAGCCTGCG CCCCCTCAGC CGGGGTGTGC ACAATGAAAT
 651 CCTGTATGTA TTCGATGCCG TCCTGCCCGA AACCTTCCTG CCTGAAATC
 701 AGGATGGCGA AGTGGCGGGT TTTGAGAAAA TGGACATCGG CGGTCTGTTG
 751 GCTGCCATGT TGTCGGGAAA CATGATGCAC GACCGCAAC TGGTTACGCT
 801 GGACGCGTTT TGCCGTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG
 851 AGTGGCTGGA CGGCATACGT TTATAG

This encodes a protein having amino acid sequence <SEQ ID 414>:

60 1 MPTVRFTESV SKHDLDALE WAKASYGAES CWKTLYLNGL PLGNLSPEWA
 51 ERVKKDWEAG CSESSDGIFL NADGWPMGR RLQHLARIWK EAGLLHGWRD

-256-

101 ECFDLTDGGS NPLFALERAA FRPFGLLSRA VHLNGLVESD GRWHFWIGRR
 151 SPHKAVDPDK LDNTAAGGVS SGELPSETVC RESSEEAGLD KTLPLIRPV
 201 SQLHSLRPVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGGLL
 251 AAMLSGNMMH DAQLVTLD AF CRYGLIDAAH PLSEWLDGIR L*

5 ORF105a and ORF105-1 show 93.8% identity in 291 aa overlap:

		10	20	30	40	50	60
	orf105a.pep	MPTVRF	TESVSKHDL	DALFEWAKASYGAESC	WKTLYL	NGLPLGNLSPEWA	ERVKKDWEAG
	orf105-1	MPTVRF	TESVSKQDL	DALFEWAKASYGAESC	WKTLYL	NGLPLGNLSPEW	ERVKKDWEAG
10		10	20	30	40	50	60
	orf105a.pep	CSESSD	GIFLNADGW	PDMGRRLOHLARI	WKEAGLLHGW	RDEC	FDLTDGGSNPLFALERAA
	orf105-1	CSESSD	GIFLNADGW	PDMGRRLOHLALG	WHCAGLLDG	WRNECFDLTDG	GGNPLFTLERAA
15		70	80	90	100	110	120
	orf105a.pep	FRPFGL	LSRAVHLNGLVES	DGRWHFWIGRR	SPHKAVDPDK	LDNTAAGGV	SSGELPSETVC
	orf105-1	FRPFGL	LSRAVHLNGLTES	DGRWHFWIGRR	SPHKAVDPN	KLDNTAAGGV	SSGEMPSEAVC
20		130	140	150	160	170	180
	orf105a.pep	RESSEE	AGLDKTLPLIR	PVSQ	LHSLRPVSR	GVHNEILYV	FDAVLPETFLPENQDGEVAG
	orf105-1	RESSEE	AGLDKTLPLIR	PVSQ	LHSLRVS	RGVHNEILYV	FDAVLPETFLPENQDGEVAG
25		190	200	210	220	230	240
	orf105a.pep	FEKMDI	GGLLAAMLSGN	MMHDAQLVT	LD AF CRYGLI	DAAHPLSEW	LDGIRLX
	orf105-1	FEKMDI	GGLLDAMLSGN	MMHDAQLVT	LD AF CRYGLI	DAAHPLSEW	LDGIRLX
30		250	260	270	280	290	
	orf105a.pep	NEILYV	FDAVLP				
	orf105-1	NEILYV	FDAVLP				
35							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF105 shows 87.5% identity over a 312aa overlap with a predicted ORF (ORF105.ng) from *N.*

gonorrhoeae:

	orf105.pep	MVARRAHNPKVVGSNPXPATXFQTPRFNAEXVLXLPVSCFLFPAASVFCRIFLPAAISER	60
	orf105ng	MVARRAHNPKVVGSNPAPATKYQTPRFNAEGVLF-----FLFPAASVFCRIFLPAAISER	55
40			
	orf105.pep	QTAVCLRLQIQAVWLQSSALSSRKPTMPTVRF	120
	orf105ng	QAAVCLRLQIQAVWLQSSALCSRKPAMPTVRF	115
45			
	orf105.pep	LYLNGXPLGNLSPEWVERVXKDWEAGCXESSD	180
	orf105ng	LYLNRLPLGNLSPEWAERIKKDWEAGCSESSNG	175
50			
	orf105.pep	LDGWRNECFDLTDGGGNPLFTLERAXRXPXG	240
	orf105ng	LHGWRNECFDLTDGGGNPLFTLERAAFRPFGL	235
55			
	orf105.pep	AVDENKLDNTXAGGVSGGEMPSEAVCRESS	300
	orf105ng	AVDPGKLDNIAGGVSGGEMPSEAVCRESS	295
	orf105.pep	NEILYVDAVLP	312
60			
	orf105ng	NEILYVDAVLPETFLPENQDGEVAGFEKMDI	355

A complete length ORF105ng nucleotide sequence <SEQ ID 415> was predicted to encode a protein having amino acid sequence <SEQ ID 416>:

-257-

1 MVARRAHNPV VVGSNPAPAT KYOTPRFNAE GVLFFLPFAA SVFCRIFLPA
 51 AISERQAQAVC LRLQIQAVWL QSSALCSRKP AMPTVRFTE VSKQDLDALE
 101 ERAKASYGAE SCWKTLYLNR LPLGNLSPEW AERIKKDWEA GCSESSNGIF
 151 LNADGWPDMD GRLQHLARTW NKAGLLHGW NRWHEWIGR RSPHKAVDPG KLDNIAGGGV
 5 S 201 AFRPFGLLIR AVHLNGLVES NGRWHEWIGR RSPHKAVDPG KLDNIAGGGV
 251 SGGEMPSEAV CRESSEAGL DKTLLFLIRP VSRHLSLRPV SRGVHNEILY
 301 VFDAVLPETF LPENQDGEVA GFEMDIGGL LDAMLSKNMM HDAQLVTLDA
 351 FYRYGLIDAA HPLSEWLDGI RL*

Further work revealed the complete nucleotide sequence <SEQ ID 417>:

10 1 ATGCCGACCG TCCGTTTAC CGAATCCGTC AGCAACAAG ACCTTGATGC
 51 CCTGTTCGAG CGGGCAAAAG CAAGTTACGG TGCCGAAAGT TGCTGGAAAA
 101 CGCTGTATCT GAACCGTCTT CCTTTGGGCA ATCTGTGCGC GGAATGGGCT
 151 GAGCGCATCA AAAAGACTG GGAGGCAGGC TGCTCCGAGT CTTGAGACGG
 201 CATTTTTCTG AATGCGGACG GCTGGCCGGA TATGGCCGGA CGCTTGACG
 15 ACCTCGCCCG CACATGGAAC AAGGCGGGG TGCTTCACGG ATGGCGCAAC
 301 GAGTGTTCG ACCTGACCGA CGGCGGCGGC AACCCCTTGT TCACGCTCGA
 351 ACGCGCCGCT TTCCGTCCGT TCGGACTACT CAGCCGCGCC GTCCATCTCA
 401 ACGGTTTGGT CGAATCGAAC GGCAGATGGC ATTTTGGAT AGGCAGGCGC
 451 AGTCCGCACA AAGCAGTCGA tCCCGCAAG CTCGACAATA TTGCCGCGCG
 20 501 CGGTGTTTCC GGCAGCGAAA TGCCGTCTGA AGCCGTGTGC CGCGAAAGCA
 551 GCGAAGAAGC CGGTTTGGAT AAAACGCTGT TTCCGCTCAT CCGCCAGTA
 601 TCGCGGCTGC ACAGCCTTCG CCCCCTCAGC CGAGGTGTGC ACAATGAAAT
 651 CCTGTATGTG TTCGATGCCG TCCTGCCCCA AACCTTCCTG CCTGAAAATC
 701 AGGATGGCGA GGTAGCGGGT TTTGAAAAGA TGGACATTGG CGGCCTATTG
 25 751 GATGCCATGT TGTCGAAAAA CATGATGCAC GACGCGCAAC TGGTTACGCT
 801 GGACGCGTTT TACCGTTACG GTCTGATTGA TGCCGCCCAT CCGCTGTCCG
 851 AGTGGCTGGA CGGCATACGT TTATAG

This corresponds to the amino acid sequence <SEQ ID 418; ORF105ng-1>:

1 MPTVRFTE SV SKQDLDALE RAKASYGAE CWKTLYLNRL PLGNLSPEWA
 30 51 ERIKKDWEAG CSESSDGIFL NADGWPDMDG RLQHLARTWN KAGLLHGWNR
 101 ECFDLTDGGG NPLFTLERAA FRPFGLLSRA VHLNGLVESN GRWHEWIGRR
 151 SPHKAVDPGK LDNIAGGGVS GGEMPSEAVC RESSEAGLD KTLFLIRPV
 201 SRLHSLRPVS RGVHNEILYV FDAVLPETFL PENQDGEVAG FEKMDIGGLL
 251 DAMLSKNMMH DAQLVTLDAF YRYGLIDAAH PLSEWLDGIR L*

35 ORG105ng-1 and ORF105-1 show 93.5% identity in 291 aa overlap:

		10	20	30	40	50	60
orf105-1.pep		MPTVRFTE SV SKQDLDALE	FEWAKASYGAE	SCWKTLYLNRL	PLGNLSPEW	VERVKKDWEAG	
40 orf105ng-1		MPTVRFTE SV SKQDLDALE	FERAKASYGAE	SCWKTLYLNRL	PLGNLSPEW	AERIKKDWEAG	
		10	20	30	40	50	60
		70	80	90	100	110	120
orf105-1.pep		CSESSDGIFL	NADGWPDMDG	RLQHLALG	WHCAGLLD	GWNECFDLTD	GGGNPLFTLERAA
45 orf105ng-1		CSESSDGIFL	NADGWPDMDG	RLQHLARTWN	KAGLLHGWNR	NECFDLTD	GGGNPLFTLERAA
		70	80	90	100	110	120
		130	140	150	160	170	180
orf105-1.pep		FRPFGLLSRA	VHLNGLTESD	GRWHEWIGRR	SPHKAVDPN	KLDNTAAGGV	SGGEMPSEAVC
50 orf105ng-1		FRPFGLLSRA	VHLNGLVESN	GRWHEWIGRR	SPHKAVDPG	KLDNIAGGGV	SGGEMPSEAVC
		130	140	150	160	170	180
		190	200	210	220	230	240
orf105-1.pep		RESSEAGLD	DKTLLPLIRP	VSQLHSLRS	VRGVHNEILY	VFVAVLPE	TFLPENQDGEVAG
55 orf105ng-1		RESSEAGLD	DKTLLPLIRP	VSRHLSLRP	VRGVHNEILY	VFVAVLPE	TFLPENQDGEVAG
		190	200	210	220	230	240
		250	260	270	280	290	
orf105-1.pep		FEKMDIGGL	LDAAMLSGNM	MHDAQLVTL	DAFCRYGLI	DAAHPLSEW	LDGIRLX
orf105ng-1		FEKMDIGGL	LDAAMLSKNM	MHDAQLVTL	DAFYRYGLI	DAAHPLSEW	LDGIRLX
		250	260	270	280	290	

Furthermore, ORF105ng-1 shows homology with a yeast enzyme:

```

5  sp|P41888|TNR3_SCHPO THIAMIN PYROPHOSPHOKINASE (TPK) (THIAMIN KINASE)
   >gi|1076928|pir|S52350 thiamin pyrophosphokinase (EC 2.7.6.2) - fission yeast
   (Schizosaccharomyces pombe) >gi|666111 (X84417) thiamin pyrophosphokinase
   (Schizosaccharomyces pombe) >gi|2330852|gnl|PID|e334056 (298533) thiamin
   pyrophosphokinase (Schizosaccharomyces pombe) Length = 569
   Score = 105 bits (259), Expect = 4e-22
   Identities = 64/192 (33%), Positives = 94/192 (48%), Gaps = 3/192 (1%)
10 Query: 268 NKAGLLHGWRNECFDLTDGGGNPLFTLERAARFPFGLLSRAVHLNGLVESNGRW--HFWI 441
   N G+ WRNE + + P+ +ER F FG LS VH + + W+
   Sbjct: 96 NTFGIADQWRNELYTVYGKSKKPVLAVERRGGFWLFGFLSTGVHCTMYIPATKEHPLRIWV 155
15 Query: 442 GRRSPHKAVDPGKLDNIAGGGVSGGEMPSEAVCRESEEAGLDKTLFPLIRPVSRLHSLR 621
   RRSP K P LDN GG++ G+ + +E SEEA LD + LI P + ++
   Sbjct: 156 PRRSPTKQTWPNYLDNSVAGGIAHGDSVIGTMIKEFSEANLDVSSMNLIPCGTVSYIK 214
20 Query: 622 PVSRG-VHNEILYVFDVAVLPETFLPENQDGEVAGFEKMDIGLLDAML SKNMHDAQLVT 798
   R + E+ YVFD + + +P DGEVAGF + + +L + K+ + LV
   Sbjct: 215 MEKRHWIQPELQYVFDLPVDDLVIPIRINDGEVAGFSLPLNQVLHELELKSFKPCALVL 274
   Query: 799 LDAFYRYGLIDAAHP 843
   LD R+G+I HP
   Sbjct: 275 LDFLIRHGIITPQHP 289

```

25 Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 49

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

30 419>:

```

1  ATGAATAGAC CCAAGCAACC CTTCTTCCGT CCCGAAGTCG CCGTTGCCCG
51  CCAAACCAGC CTGACGGGTA AAGTGATTCT GACACGACCG TTGTCATTTT
101 CCCTATGGAC GACATTTGCA TCGATATCTG CGTTATTGAT TATCCTGTTT
151 TTGATATTG GTAACATATC GCGAAAGACA ACAGTGGAGG GACAAATTTT
35 201 ACCTGCATCG GGCCTAATCA GGGTGTATGC ACCGgATACG rGkACAATTA
251 CAGCGAAATT CGTGGAAAGT GgmsAAAAGG TTAAGGCTGG CGACAAGCTA
301 TTTGCGCTTT CGACCTCAGC TTTCGGCGCA GGAGGTAGCG TGCAGCAGCA
351 GTTGAACACG GAGGCAGTTT TGAAGAAAAC GTTGGCAGAA CAGGAACCTG
401 GTCGTCTGAA GCTGATACAC GGAATGAAA CGCGCAGCct TAAAGCAACT
40 451 GTCGAACGTT TGGAAAACCA GGAACCTCCAT ATTTGCAAC AGATAGACGG
501 TCAGAAAAGG CGCATTAGAC TTGCGGAAGA AATGTTGCAG AAATATCGTT
551 TCCTATCCGC .CAATGA

```

This corresponds to the amino acid sequence <SEQ ID 420; ORF107>:

```

45 1  MNRPKQFFER PEVAVARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
51  LIFGNYTRKT TVEGQILPAS GVIRVYAPDT XTITAKFVED GXKVKAGDKL
101 FALSTSRFGA GGSVQQQLKT EAVLKKT LAE QELGRLKLIH GNETRSLKAT
151 VERLENQELH ISQQIDGQKR RIRLAEMLQ KYRFLSXQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF107 shows 97.8% identity over a 186aa overlap with an ORF (ORF107a) from strain A of *N.meningitidis*:

10 20 30 40 50 60

-259-

5	orf107.pep	MNRPKQFFRPEVAVARQTSLTGKVILTRPLSFSLWTTFASISALLIILFLIFGNYTRKT	
	orf107a	MNRPKQFFRPEVAVARQTSLTGKVILTRPLSFSLWTTFASISALLIILFLIFGNYTRKT	10 20 30 40 50 60
10	orf107.pep	TVEGQILPASGVIRVYAPDXTITAKFVEDGXKVKAGDKLFALSTSRFGAGGSVQQQLKT	70 80 90 100 110 120
	orf107a	TVEGQILPASGVIRVYAPDGTITAKFXEDGEKVKAGDKLFALSTSRFGAGGSVQQQLKT	70 80 90 100 110 120
15	orf107.pep	EAVLKKTAEQELGRLKLIHNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLO	130 140 150 160 170 180
	orf107a	EAVLKKTAEQELGRLKLIHNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLO	130 140 150 160 170 180
20	orf107.pep	KYRFLSXQX	189
	orf107a	KYRFLSANDAVPKQEMNVKAEELLEQKAKLDAYRREEVGLLQEIQTQNLTLXSLPQAAAX	190 200 210 220 230

The complete length ORF107a nucleotide sequence <SEQ ID 421> is:

25	1	ATGAATAGAC	CCAAGCAACC	NTTCTTCCGT	CCCGAAGTCG	CCGTTGCCCG
	51	CCAAACCAGC	CTGACGGGTA	AAGTGATTCT	GACACGACCG	TTGTCATTTT
	101	CCCTATGGAC	GACATTTGCA	TCGATATCTG	CGTTATTGAT	TATCCTGTTT
	151	TTGATATTG	GTAACATAC	GCGAAAGACA	ACAGTGGAGG	GACAAATTTT
	201	ACCTGCATCG	GGCGTAATCA	GGGTGTATGC	ACCGGATACG	GGGACAATTA
	251	CNGCGAAATT	CNTGGAAGAT	GGAGAAAAGG	TTAAGGCTGG	CGACAAGCTA
30	301	TTGCGCTTT	CGACCTCAG	TTTCGGCGCA	GGAGATAGCG	TGCAGCAGCA
	351	GTTGAAACG	GAGGCAGTTT	TGAAGAAAAC	GTTGGCAGAA	CAGGAACCTGG
	401	GTCGTCTGAA	GCTGATACAC	GGGAATGAAA	CGCGCAGCCT	TAAAGCAACT
	451	GTCGAACGTT	TGGAAAACCA	GGAATCCAT	ATTTCGCAAC	AGATAGACGG
	501	TCAGAAAAGG	CGCATTAGAC	TTGCGGAAGA	AATGTGCAG	AAATATCGTT
35	551	TCCTATCCGC	CAATGATGCA	GTGCCAAAAC	AAGAAATGAT	GAATGTCAAG
	601	GCAGAGCTTT	TAGAGCAGAA	AGCCAAACTT	GATGCCTACC	GCCGAGAAGA
	651	AGTCGGGCTG	CTTCAGGAAA	TCCGCACGCA	GAATCTGACA	TTGGNNAGCC
	701	TCCCCAAGC	GGCATGA			

This encodes a protein having amino acid sequence <SEQ ID 422>:

40	1	MNRPKQFFR	PEVAVARQTS	LTGKVILTRP	LSFSLWTTFA	SISALLIILF
	51	LIFGNYTRKT	TVEGQILPAS	GVIRVYAPDT	GTITAKFXED	GEKVKAGDKL
	101	FALSTSRFGA	GDSVQQQLKT	EAVLKKTAE	QELGRLKLIH	GNETRSLKAT
	151	VERLENQELH	ISQQIDGQKR	RIRLAEEMLO	KYRFLSANDA	VPKQEMNVK
45	201	AELLEQKAKL	DAYRREEVGL	LQEIRTQNL	LXSLPQAA*	

Homology with a predicted ORF from *N. gonorrhoeae*

ORF107 shows 95.7% identity over a 188aa overlap with a predicted ORF (ORF107.ng) from *N. gonorrhoeae*:

50	orf107.pep	MNRPKQFFRPEVAVARQTSLTGKVILTRPLSFSLWTTFASISALLIILFLIFGNYTRKT	60
	orf107ng	MNRPKQFFRPEVAIARQTSLTGKVILTRPLSFSLWTTFASISALLIILFLIFGNYTRKT	60
	orf107.pep	TVEGQILPASGVIRVYAPDXTITAKFVEDGXKVKAGDKLFALSTSRFGAGGSVQQQLKT	120
55	orf107ng	TMEGQILPASGVIRVYAPDGTITAKFVEDGEKVKAGDKLFALSTSRFGAGGSVQQQLKT	120
	orf107.pep	EAVLKKTAEQELGRLKLIHNETRSLKATVERLENQELHISQQIDGQKRRIRLAEEMLO	180
	orf107ng	EAVLKKTAEQELGRLKLIHNETRSLKATVERLENQKLHISQQIDGQKRRIRLAEEMLR	180
60	orf107.pep	KYRFLSXQ 188	
	orf107ng	KYRFLSAQ 188	

The complete length ORF107ng nucleotide sequence <SEQ ID 423> is predicted to encode a protein having amino acid sequence <SEQ ID 424>:

```

1  MNRPKQPFRR PEVAIARQTS LTGKVILTRP LSFSLWTTFA SISALLIILF
5  51  LIFGNVTRKT TMEGQILPAS GVIRVYAPDT GTITAKFVED GEKVKAGDKL
101 FALSTSRFGA GGSVQQQLKT EAVLKKTLAE QELGRLKLIH ENETRSKAT
151 VERLENQKLH ISQQIDGQKR RIRLAEEMLR KYRFLSAQ*

```

Based on the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 50

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 425>:

```

1  ATGCTGAATA CTTTTTTTGC CGTATTGGGC GGCTGCCTGC TGCT.TTGCC
15 51  GTGCGGCAAA TCCGTAAATA CGCGGTGACA GCCGCAAAAC GCGGTACAAA
101 GCGCGCCGAA ACCGGTTTTC AAAGTCATAT ATATCGACAA TACGGCGATT
151 GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA
201 AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC
251 GACTGATCGG CAAGCATCCC GCGCACTTGG AAGCCGTCAG CGGCAAATGT
301 ATGGAAACCG ATGATAAGGA CAGTCCGGCA GGTGGGCGAG AAAACGGCGT
20 351 GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
401 GCAAACGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG
451 GCAGGCAAAA GCGGCTATGC CGCGTGCGAG AACGGACGCT ATGTGCTGGA
501 AATCGACAGC GAAGGGGCGT TTTATTTCCG CCGCCGCCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 426; ORF108>:

```

25 1  MLNTFFAVLG GCLLXLPCGK SVNTAVQPQN AVQSAPKPVF KVIYIDNTAI
51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
101 METDDKDSPA GWAENGVCHT LFAKLVGNIA EDGGKLTDL VSHAALQPYQ
151 AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

Further work revealed the following DNA sequence <SEQ ID 427>:

```

30 1  ATGCTGAAAA CATCTTTTGC CGTATTGGGC GGCTGCCTGC TGCTTGCCCGC
51  CTGCGGCAAA TCCGAAAATA CGGCGGAACA GCCGCAAAAC GCGGTACAAA
101 GCGCGCCGAA ACCGGTTTTC AAAGTCAAAT ATATCGACAA TACGGCGATT
151 GCCGGTTTGG ATTTGGGACA AAGCAGCGAA GGCAAAACCA ACGACGGCAA
35 201 AAAACAAATC AGTTATCCGA TTAAAGGCTT GCCGGAACAA AATGTTATCC
251 GACTGATCGG CAAGCATCCC GCGCACTTGG AAGCCGTCAG CGGCAAATGT
301 ATGGAAACCG ATGATAAGGA CAGTCCGGCA GGTGGGCGAG AAAACGGCGT
351 GTGCCATACC TTGTTTGCCA AACTGGTGGG CAATATCGCC GAAGACGGCG
401 GCAAACGAC GGATTACCTA GTTTCGCATG CCGCCCTGCA ACCCTATCAG
451 GCAGGCAAAA GCGGCTATGC CGCGTGCGAG AACGGACGCT ATGTGCTGGA
40 501 AATCGACAGC GAAGGGGCGT TTTATTTCCG CCGCCGCCAT TATTGA

```

This corresponds to the amino acid sequence <SEQ ID 428; ORF108-1>:

```

45 1  MLKTSFAVLG GCLLLAACGK SENTAEQPQN AVQSAPKPVF KVKYIDNTAI
51  AGLDLGQSSE GKTNDGKKQI SYPIKGLPEQ NVIRLIGKHP GDLEAVSGKC
101 METDDKDSPA GWAENGVCHT LFAKLVGNIA EDGGKLTDL VSHAALQPYQ
151 AGKSGYAAVQ NGRYVLEIDS EGAFYFRRRH Y*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF108 shows 88.4% identity over a 181aa overlap with a predicted ORF (ORF108.ng) from *N.gonorrhoeae*:

```

5      orf108.pep  MLNTFFAVLGGCLLXLPCGKSVNTAVQPQNAVQSAPKPVFKVIYIDNTAIAGLDLGQSSE  60
      orf108ng    MLKIPFAVLGGCLLLAACGKSENTAEQPQNAQAQSAKPVFKVKYIDNTAIAGLALGQSSE  60

10     orf108.pep  GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT  120
      orf108ng    GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT  120

15     orf108.pep  LFAKLVGNIAEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH  181
      orf108ng    LFAKLVGNIAEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH  181

```

ORF108-1 shows 92.3% identity with ORF108ng over the same 181 aa overlap:

```

20     orf108-1.pep MLKTSFAVLGGCLLLAACGKSENTAEQPQNAVQSAPKPVFKVKYIDNTAIAGLDLGQSSE  60
      orf108ng-1    MLKIPFAVLGGCLLLAACGKSENTAEQPQNAQAQSAKPVFKVKYIDNTAIAGLALGQSSE  60

      orf108-1.pep  GKTNDGKKQISYPIKGLPEQNVIRLIGKHPGDLEAVSGKCMETDDKDSPAGWAENGVCHT  120
      orf108ng-1    GKTNDGKKQISYPIKGLPEQNAVRLTGKHPNDLEAVVGKCMETDGKDAPSGWAENGVCHT  120

25     orf108-1.pep LFAKLVGNIAEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH  181
      orf108ng-1    LFAKLVGNIAEDGGKLT DYLVSHAALQPYQAGKSGYAAVQNGRYVLEIDSEGAFFRRRH  181

```

The complete length ORF108ng nucleotide sequence <SEQ ID 429> is:

```

30      1  ATGCTGAAAa  tacctTTTGC  CGTGTtgggc  ggCtgccctGC  TGCTTGCCGC
      51  CTGCGGCAAA  TCCGAAAATa  cggcggaACA  GCCGCAAAAT  gcggCACAAA
101     GCGCGCCGAA  ACCGGTTTTC  AAAGTCAAAT  ACATCGACAA  TACGGCGATT
      151  GCCGGTTTGG  CTTTGGGACA  AAGTAGCGAA  GGCAAAACCA  acgacgGCAA
      201  AAAACAAATC  AGTTATccgA  TTAAAGGCTT  GCCGGAACAA  Aacgcgctcc
      251  gGCTGACCGG  AAAGCATCCC  AACGACTTGG  AagccgtcgT  CGGCAAATGT
35      301  ATGGAACCG  ACGGAAAGGA  CGCGCCTTCG  GGCTGGGCGG  AAAACGGCGT
      351  GTGCCATACC  TTGTTTGCCA  AACTGGTGGG  CAATATCGCC  GAAGACGGCG
      401  GCAAACTGAC  TGATTACCTG  ATTTTCGATT  CCGCCCTGCA  ACCCTATCAG
      451  GCAGGCAAAA  GCGGCTATGC  CGCCGTGCAG  AACGGACGCT  ATGTGCTGGA
      501  AATCGACAGC  GagggGGCGT  TTTATttccg  ccgcccctat  tattgA

```

40 This encodes a protein having amino acid sequence <SEQ ID 430>:

```

      1  MLKIPFAVLG  GCLLLAACGK  SENTAEQPON  AAQSAKPVF  KVKYIDNTAI
      51  AGLALGQSSE  GKTNDGKKQI  SYPIKGLPEQ  NAVRLTGKHP  NDLEAVVGKC
      101  METDGKDAPS  GWAENGVCHT  LFAKLVGNIA  EDGGKLT DYL  ISHSALQPYQ
      151  AGKSGYAAVQ  NGRYVLEIDS  EGAFYFRRRH  Y*

```

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) and a putative ATP/GTP-binding site motif A (P-loop, double-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

50 Example 51

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 431>:

1 ATGGAAGATT TATATATAAT ACTCGCTTTG GGTTTGGTTG CGATGATTGC
 51 CGGATTTATC GATgcatTg cGggCGGGGG TGGTTTGATT ACGCTGCCCCG
 101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
 151 CTGCAAGCAG CCGCTGCTAC GTTTTCAGCT ACGGTTTCTT TTGCACGCAA
 201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
 251 TAGGCGGCGT GGcCGGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
 301 CTgCTgGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGAC TGTATTTTGT
 351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
 401 TTTTCTGTT cGGGCTGACG GTCGC.ACCG CTTTGGGTT TTTACGACGG
 451 TGTGTTCCGA CCGGTGTGCG GCTCGTTTTT TCTGATTGCC TTTATTGTTT
 501 TGCTCGGCTG CAAgCTGTTG AACGCGATGT CTTACACCAA ATTGGCGAAC
 551 GTTGCCCTGCA ATCTTGGTTC GCTATCGGTA TTCCTGCTGC ACGGTTTCGAT
 601 TATTTTCCCG ATTGCGGCAA CGaTGCGGT CCGTGCGTTT GTCGGTGGCA
 651 ATTTAGTGTC GAGATTGACC GTaCgctTCG GTTCGAAGCT GATTAA

15 This corresponds to the amino acid sequence <SEQ ID 432; ORF109>:

1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
 51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI
 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VXTAFGLRR
 151 CVRTGCRLVF SDCLYCFARL QAVRDVLHQ IGERCLQSWF AIGIPAARFD
 201 YFPDCGNDGG RCVCRCEERC EICRTLFEA D*

Further work revealed the following DNA sequence <SEQ ID 433>:

1 ATGGAAGATT TATATATAAT ACTCGCTTTG GGTTTGGTTG CGATGATTGC
 51 CGGATTTATC GATGCGATTG CCGGCGGGGG TGGTTTGATT ACGCTGCCCCG
 101 CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
 151 CTGCAAGCAG CCGCTGCTAC GTTTTCAGCT ACGGTTTCTT TTGCACGCAA
 201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
 251 TAGGCGGCGT GGcCGGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
 301 CTGCTGCGG TCGTGCCGGT TTTGTTGATA TTTGTGCGAC TGTATTTTGT
 351 GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
 401 TTTTCTGTT CCGGCTGACG GTCGCACCGC TTTTGGGTTT TTACGACGGT
 451 GTGTTCCGAC CCGGTGTGCG CTCGTTTTT CTGATTGCCT TTATTGTTTT
 501 GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAA TTGGCGAACG
 551 TTGCCTGCAA TCTTGGTTCG CTATCGGTAT TCCTGCTGCA CCGTTCGATT
 601 ATTTTCCCGA TTGCGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
 651 TTTAGGTGCG AGATTTGCCG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
 701 TGCTGATTGT CATCAGCATT TCGATGGCTG TGAATTTGTT GATAGACGAG
 751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTTTAA

This corresponds to the amino acid sequence <SEQ ID 434; ORF109-1>:

1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
 51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFVGGVAGA LSVSLVSKDI
 101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
 151 VFGPGVGSFF LIAFIVLLGC KLLNAMSYTK LANVACNLGS LSVFLLHGS
 201 IFPIAATMAV GAFVGANLGA RFAVRFGSKL IKPLLVISI SMAVKLLIDE
 251 RNPLYQMIVS MF*

45 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF109 shows 95.9% identity over a 147aa overlap with an ORF (ORF109a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
50	orf109.pep	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNK	LQAAAATFSA				
	orf109a	MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNK	LQAAAATFSA				
		10	20	30	40	50	60
55	orf109.pep	TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
	orf109a	TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP					
		70	80	90	100	110	120

```

      130      140      150      160      170      180
orf109.pep KLDGSKEGKARMSFFLFGLTVXTAFGFLRRCVRTGCRVFSDCLYCFARLQAVRDVLHQ
            |||||
5 orf109a   KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMSYTK
            130      140      150      160      170      180

```

The complete length ORF109a nucleotide sequence <SEQ ID 435> is:

```

1 ATGGAAGATT TATACATAAT ACTCGCTTTG GGTTCGTTG CGATGATTGC
51 CCGATTATC GATGCGATTG CCGGTGGGGG TGGTTTGATT ACGCTGCCTG
10 101 CACTCTTGT GGCAGGTATT CCTCCGTGT CCGCAATTGC CACCAACAAG
151 CTGCAAGCAG CCGCTGCTAC GTTTTCGGCT ACGGTTTCTT TTGCACGCAA
201 AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCGGCA GCATCGTTTG
251 CAGGCGGCGT GGTGCGTGCA TTATCGGTCA GCTTGGTTTC CAAAGATATT
301 CTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTGTGCGCGC TGTATTTGT
15 351 GTTTTCGCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
401 TTTTCTGTT CCGTCTGACG GTTGCAACAC TTTTGGGTTT TTACGACGGT
451 GTGTTGCGAC CCGGTGTCGG CTCGTTTTTT CTGATTGCCT TTATTGTTT
501 GCTCGGCTGC AAGCTGTGA ACGCGATGTC TTACACCAAA TTGGCGAACG
551 TTGCCTGCAA TCTTGTTTCG CTATCGGTAT TCCTGCTGCA CGGTTTCGATT
20 601 ATTTCCCGA TTGCGGCAAC GATGGCGGTC GGTGCGTTTG TCGGTGCGAA
651 TTTAGGTGCG AGATTGCGG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
701 TGCTGATTGT CATCAGCATT TCGATGGCTG TGAATTGTT GATAGACGAG
751 AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTAA

```

This encodes a protein having amino acid sequence <SEQ ID 436>:

```

25 1 MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
101 LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
151 VFGPGVGSFF LIAFIVLLGC KLLNAMSYTK LANVACNLGS LSVFLLHGS
201 IFPIAATMAV GAFVGANLGA RFAVRFGSKL IKPLLIVISI SMAVKLLIDE
30 251 RNPLYQMIVS MF*

```

ORF109a and ORF109-1 show 99.2% identity in 262 aa overlap:

```

      10      20      30      40      50      60
orf109a.pep MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
35 orf109-1 MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
      10      20      30      40      50      60

      70      80      90      100     110     120
orf109a.pep TVSFARKGLIDWKKGLPIAAASFAGGVVGA LSVSLVSKDILLAVVPVLLIFVALYFVFSP
40 orf109-1 TVSFARKGLIDWKKGLPIAAASFAGGVVGA LSVSLVSKDILLAVVPVLLIFVALYFVFSP
      70      80      90      100     110     120

      130     140     150     160     170     180
orf109a.pep KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMSYTK
45 orf109-1 KLDGSKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFFLIAFIVLLGCKLLNAMSYTK
      130     140     150     160     170     180

      190     200     210     220     230     240
orf109a.pep LANVACNLGSLSVFLLHGSIIIFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
50 orf109-1 LANVACNLGSLSVFLLHGSIIIFPIAATMAVGAFVGANLGARFAVRFGSKLIKPLLIVISI
      190     200     210     220     230     240

      250     260
orf109a.pep SMAVKLLIDERNPLYQMIVSMFX
55 orf109-1 SMAVKLLIDERNPLYQMIVSMFX
      250     260
60

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF109 shows 98.3% identity over a 231aa overlap with a predicted ORF (ORF109.ng) from *N. gonorrhoeae*:

```

5      orf109.pep  MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA  60
      orf109ng    MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA  60

10     orf109.pep  TVSFARKGLIDWKKGLPIAAASFVGGVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP  120
      orf109ng    TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP  120

15     orf109.pep  KLDGSKEGKARMSFFLFGLTVXTAFGFLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ  180
      orf109ng    KLDGSKEGKARMSFFLFGLTVATAFGFLRRCVRTGCRLVFSDCLYCFARLQAVERDVLHQ  180

      orf109.pep  IGERCQSWFAIGIPAARFDYFPDCGNDGGRCVCRCEFRCEICRPLRFEAD  231
      orf109ng    IGERCQSWFAIGIPAARFDYFPDCGNDGGRCVCRCEFRCEICRPLRFEAD  231

```

An ORF109ng nucleotide sequence <SEQ ID 437> was predicted to encode a protein having amino acid sequence <SEQ ID 438>:

```

1      MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
51     LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
101    LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VATAFGFLRR
151    CVRTGCRLVF SDCLYCFARL QAVERDVLHQ IGERCQSWF AIGIPAARFD
25     201 YFPDCGNDGG RCVCRCEFRCE EICRPLRFEA D*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 439>:

```

1      ATGGAAGATT TATACATAAT ACTCGCTTTG GGTTCGGTTG CGATGATCGC
51     CGGATTATC  GATGCGATTG CCGGCGGGGG TGGTTTGATT ACGCTGCCTG
101    CACTCTTGTT GGCAGGTATT CCTCCCGTGT CGGCAATTGC CACCAACAAG
151    CTGCAAGCAG CCGCTGCTAC GTTTTCGGCT ACGGTTTCTT TTGCACGCAA
201    AGGTTTGATT GATTGGAAGA AAGGTCTCCC GATTGCCGCA GCATCGTTTG
251    CAGGCGCGGT GGTCGGTGCA TTATCGGTCA GCTTGCTTTC CAAAGATATT
301    TTGCTGGCGG TCGTGCCGGT TTTGTTGATA TTTGTCGCGC TGTATTTTGT
351    GTTTTCGCCC AAGCTCGACG GCAGTAAGGA AGGCAAAGCC AGAATGTCTT
401    TTTTCTTATT CGGGCTGACG GTTGCAACGC TTTTGGGTTT TTACGACGGT
451    GTGTTTCGGAC CGGGTGTCGG CTCGTTTTTT CTGATTGCCT TTATTGTTTT
501    GCTCGGCTGC AAGCTGTTGA ACGCGATGTC TTACACCAA TTGGCGAAGC
551    TTGCTTGCAA TCTTGTTTCG CTATCGGTAT TCCTGCTGCA CGGTTTCGATT
601    ATTTTCCCGA TTGTGGCAAC GATGCGGTC GGTGCGTTTG TCGGTGCGAA
651    TTTAGGTGCG AGATTGCGG TCCGCTTCGG TTCGAAGCTG ATTAAGCCGC
701    TGCTGATTGT CATCAGCATT TCGATGGCTG TGAATTGTT GATAGACGAG
751    AGAAATCCGC TGTATCAGAT GATTGTTTCG ATGTTTAA

```

This corresponds to the amino acid sequence <SEQ ID 440; ORF109ng-1>:

```

1      MEDLYIILAL GLVAMIAGFI DAIAGGGGLI TLPALLLAGI PPVSAIATNK
45     51 LQAAAATFSA TVSFARKGLI DWKKGLPIAA ASFAGGVVGA LSVSLVSKDI
101    LLAVVPVLLI FVALYFVFSP KLDGSKEGKA RMSFFLFGLT VAPLLGFYDG
151    VFGPGVGSFF LIAFIVLLGC KLLNAMS YTK LANVACNLGS LSVFLHGS I
201    IFPIVATMAV GAFVGANLGA RFAVRFGSKL IKPLLVISI SMAVKLLIDE
251    RNPLYQMIVS MF*

```

ORF109ng-1 and ORF109-1 show 98.9% identity in 262 aa overlap:

```

10      20      30      40      50      60
orf109ng-1.pep MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
55     orf109-1    MEDLYIILALGLVAMIAGFIDAIAGGGGLITLPALLLAGIPPVSAIATNKLQAAAATFSA
10      20      30      40      50      60

70      80      90      100     110     120
orf109ng-1.pep TVSFARKGLIDWKKGLPIAAASFAGGVVAGALSVSLVSKDILLAVVPVLLIFVALYFVFSP

```

[illegible]

In addition, ORF109ng-1 shows homology to a hypothetical *Pseudomonas* protein:

```

sp|P29942|YCB9_PSEUDE HYPOTHETICAL 27.4 KD PROTEIN IN COBO 3'REGION (ORF9)
>gi|94984|pir|I138164 hypothetical protein 9 - Pseudomonas sp >gi|551929
(M62866) ORF9 [Pseudomonas denitrificans] Length = 261
Score = 175 bits (439), Expect = 3e-43
Identities = 83/214 (38%), Positives = 131/214 (60%), Gaps = 1/214 (0%)

Query: 41 PPVSAIATNKLQXXXXXXXXXXXXXXXXXKGLIDWKKGLPIXXXXXXXXXXXXXXXXXXXXKI 100
          PP+ + TNKLQ R+G ++ K+ LP+ D+
Sbjct: 43 PPLQTLGTNKLQGLFGSGSATLSYARRGHVNLKEQLPMALMSAAGAVLGALLATIVPGDV 102

Query: 101 LLAVVPVLLIFVALYFVFSPKLDGSGKEGKARMSFFLFGLTVAPLLGFYDGVFGPGVGSFF 160
          L A++P LLI +ALYF P + G + +R++ F+F LT+ PL+GFYDGVFGPG GSFF
Sbjct: 103 LKAILPFLLLIAIALYFGLKPNM-GDVDQHSRVTPFVFTLTIVPLIGFYDGVFGPGTGSFF 161

Query: 161 LIAFIVLLGCKLLNAMSYTKLANVACNLGSLSVFLHGSIIFFPIVATMAVGAFVGANLGA 220
          ++ F+ L G +L A ++TK N N+G+ VFL G++++ + M +G F+GA +G+
Sbjct: 162 MLGFTVLTAGFVGLKATAHTKFLNFGSNVGAFGVFLFFGAVLWKVGLLMGLGQFLGAQVGS 221

Query: 221 RFAVRFGSKLIKPLLLIVISISMAVKLLIDERNPL 254
          R+A+ G+K+IKPLL+++SI++A++LL D +PL
Sbjct: 222 RYAMAKAKI IKPLLIVISIALAIRLLADPTHTPL 255

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 52

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 441>:

50	1	..CTGCTAGGGT	ATTGATCGG	TTATCGGTAC	GGCTGTTGCA	GCAAAACCAG
	51	CCGCAGACGG	ATTATTTGGT	CAAATTCGGA	TCGTTTGGG	CGAG.ATTTT
	101	TGGTTTCTG	GGACTGTATG	ACGCTATGC	TTCGGCATGG	TTTGTCTGTA
	151	TCATGATGT	TTTGGTGGTT	TCTACCAGTT	TGTGCCTGAT	TCGCAATGTG
	201	CCGCCGTTCT	GCGCGCAAAAT	GAAGTCTTTT	CGGGA AAAAG	TTAAAGAAAA
55	251	ATCTCTGGCG	GCGATGCGCC	ATTCTTCGCT	GTTGGATGTA	AAAATTGCGC
	301	CCGAGGTTGC	CAAACGTTAT	CTGGAAGTAC	AAGGTTTTC A	GGGGAAAAACC
	351	ATTAACCGTG	AAGACGGGTC	GTTTCTGATT	GCCGCCAAAA	AAGGCACAAT
	401	GAACAAATGG	GGCTATATCT	TTGCCCATGT	TGCTTTGATT	GTCATTTGCC
	451	TGGGCGGGTT	GATAGACAGT	AACCTGCTGT	TGAAACTGGG	TATGCTGACC
60	501	GGTCGGATTG	TTCCGGACAA	TCAGGCGGTT	TATGCCAAGG	ATTTC.AAGC

-266-

551 CCGAAAGTAT .TTTGGGTGC GTCCAATCTC TCATTAGGG GCAACGTCAA
601 TATTTCCG.A GGGGCAGaT GCGGATGTGG TTTTCCTGA

This corresponds to the amino acid sequence <SEQ ID 442; ORF110>:

5 1 ..LLGIASVIGT LLQONQPOTD YLVKFGSFWA XIFGFLGLYD VYASAWFVVI
 51 MMFLVVSTSL CLIRNVPPFW REMKSFREKV KEKSLAAMRH SSLLDVKIAP
 101 EVAKRYLEVQ GFQGTINRE DGSVLIAAKK GTMNKGYIF AHVALIVICL
 151 GGLIDSNLLL KLGMLTGRIF RTIRRFMPRI XKPESXFGCV QSLI*GQRQY
 201 FXRGRVRMWF S*

Computer analysis of this amino acid sequence gave the following results:

10 Homology with ORF88a from *N.meningitidis* (strain A)

ORF110 shows 91.5% identity over a 188aa overlap with ORF88a from strain A of *N. meningitidis*:

		10	20	30	40	50	60
	orf88a.pep	MSKSRSPPLLSRPWFAFFSSMRFAVALLS	LLGIASVIGT	TVLQONQPOTD	YLVKFGSFWA		
15	orf110		LLGIASVIGT	TVLQONQPOTD	YLVKFGSFWA		
			10	20	30		
		70	80	90	100	110	120
20	orf88a.pep	QIFGFLGLYDVYASAWFVIMMFLVVSTSL	CLIRNVPPFW	REMKSFREKV	KEKSLAAMRH		
	orf110	XIFGFLGLYDVYASAWFVIMMFLVVSTSL	CLIRNVPPFW	REMKSFREKV	KEKSLAAMRH		
		40	50	60	70	80	90
25	orf88a.pep	SSLLDVKIAPEVAKRYLEVQGFQGTINRED	GSVLIAAKK	GTMNKGYIF	FAHVALIVICL		
	orf110	SSLLDVKIAPEVAKRYLEVQGFQGTINRED	GSVLIAAKK	GTMNKGYIF	FAHVALIVICL		
		100	110	120	130	140	150
30	orf88a.pep	GGLIDSNLLKLKMLTGRIVPDNQAVYAKDF	KPESILGAS	NLSFRGNVNI	SEGSADVVF		
	orf110	GGLIDSNLLKLKMLTGRIFRTIRRFMPRI	XKPESXFGCV	QSLIXGQRQY	FXRGRVRMWF		
35		160	170	180	190	200	210
	orf88a.pep	LNADNGILVQDLPFEVKLKKFHIDFYNT	GMPRDFAS	DIETDKAT	GEKLERTIR	VNHPLT	
	orf110	SX					

40 However, ORF88 and ORF110 do not align, because they represent two different fragments of the same protein.

Homology with a predicted ORF from *N.gonorrhoeae*

ORF110 shows 88.6% identity over a 211aa overlap with a predicted ORF (ORF110.ng) from *N. gonorrhoeae*:

45	orf110.pep	LLGIASVIGT	TVLQONQPOTD	YLVKFGSFWA	30
	orf110.ng	MSKSRISPTLLSRPWFAFFSSMRFAVALLS	LLGIASVIGT	TVLQONQPOTD	YLVKFGPFWT 60
50	orf110.pep	XIFGFLGLYDVYASAWFVIMMFLVVSTSL	CLIRNVPPFW	REMKSFREKV	KEKSLAAMRH 90
	orf110.ng	RIFDFLGLYDVYASAWFVIMMFLVVSTSL	CLIRNVPPFW	REMKSFREKV	KEKSLAAMRH 120
	orf110.pep	SSLLDVKIAPEVAKRYLEVQGFQGTINRED	GSVLIAAKK	GTMNKGYIF	FAHVALIVICL 150
55	orf110.ng	SSLLDVKIAPEVAKRYLEVQGFQGTINRED	GSVLIAAKK	GTMNKGYIX	AHVALIVICL 180

-267-

```

orf110.pep  GGLIDSNLLKLGMLTGRIFRTIRRFMPRIKKPESXFGCVQSLIXGQRQYFXRGRVRMWF  210
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf110ng    GRLINXNLLKLGMLAGSIFRNNRRVMPRIKSPESIWGGVQSLIKGQRQYFQRGKVRMWF  240

5  orf110.pep  S  211
   |
   orf110ng   S  241

```

The complete length ORF110ng nucleotide sequence <SEQ ID 443> is predicted to encode a protein having amino acid sequence <SEQ ID 444>:

```

10      1  MSKSRIPTL LSRPWFAFFS SMRFAVALLS LLGIASVIGT VLQONQPOTD
      51  YLVKFGPFWT RIFDFLGLYD VYASAWFVVI MMFLVVTSL CLIRNVPPFW
     101  REMKSFREKV KEKSLAAMRH SSLLDVKIAP EVAKRYLEVR GFQGKTVSRE
     151  DGSVLIAAKK GTMNKWGYIX AHVALIVICL GRLINXNLLL KLGMLAGSIF
     201  RNNRRVMPRI SKPESIWGGV QSLIKGQRQY FQRGKVRMWF S*

```

- 15 Based on the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 53

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 445>:

```

20      1  ATGCCGTCTG AACACGCCT GCCGAAC TTT ATCCGCGTCT TGATATTTGC
      51  CCTGGGTTTC ATCTTCCTGA ACGCCTGTT C GGAACAAACC GCGCAAACCG
     101  TTACCCTGCA AGCGGAAACG ATGGGCACGA CCTATACCGT CAAATACCTT
     151  TCAAATAATC GGGACAAACT CCCCTCACCT GCCGAAATAC AAAAACGCAT
     201  CGATGACGCG CTTAAAGAAG TCAACCGGCA GATGTCCACC TATCAGCCCG
     251  ACTCCGAAAT CAGCCGGTTC AACCAACACA CAGCCGGCAA GCCCCCTCCG
     301  ATTTCAAGCG ACTTCGCACA CGTTACTGCC GAAGCCGTCC GCCTGAACCG
     351  CCTGACACAC GCGCGCTGG ACGTAACCGT CCGCCCCCTG GTCAACCTTT
     401  GGGGATTCGG CCCCGACAAA TCCGTTACCC GTGAACCGTC GCCGGAACAA
     451  ATCAAACAGG CGGCATCTTA TACGGGCATA GACAAAATCA TTTGAAACA
     501  AGGCAAAGAT TACGCTTCCT TGAGCAAAAC CCACCCCAAG GCCTATTTGG
     551  ATTTATCTTC GATTGCCAAA GGCTTCGGCG TTGATAAAGT TCGGGGCGAA
     601  CTGGAAAAAT ACGGCATTCA AAATTATCTG GTCGAAATCG GCGGCGAGTT
     651  GCACGGCAA AAGCAAAACG CGCGCGGCGA ACCGTGGCGC ATCGGTATCG
     701  AGCAGCCCAA TATCGTCCAA GCGGCAATA CGCAGATTAT CGTCCCGCTG
     751  AACAAACGTT CGTTGCCAC TTCCGGCGAT TACCGTATTT TCCACGTCGA
     801  TAAAAACGGC AAACGCCTCT CCCATATCAT CAACCCGAAC AACAAACGAC
     851  CCATCAGCCA CAACCTCGCC TCCATCAGCG TGGTCGCAGA CAGTGCATG
     901  ACGGCGGACG GCTTGTCAC AGGATTATTC GTATTGGGCG AAACCGAAGC
     951  CTTAAAGCTG GCAGAGCGCG AAAAATCGC TGTTCCTCTG ATTGTCAGGG
    1001  ATAAAGCGG CTACCGCACC GCCATGTCTT CCGAATTGTA AAAACTGCTC
    1051  CGCTAA

```

This corresponds to the amino acid sequence <SEQ ID 446; ORF111>:

```

45      1  MPSETRLNPF IRVLIFALGF IFLNACSEQT AQTVTLQGET MGTITYTVKYL
      51  SNNRDKLPSP AEIQKRIDDA LKEVNRQMS YQPDSEISRF NQHTAGKPLR
     101  ISSDFAHVTA EAVRLNRLTH GALDVTGVL VNLWGFPGDK SVTREPSPQ
     151  IKQAASYTGI DKIIKQKGD YASLSKTHPK AYLDLSSIAK GFGVDKVAGE
     201  LEKYGIQNYL VEIGGELHGK GKNARGEPR IGIEQPNIVQ GGNTQIIIVPL
     251  NNRSLATSGD YRIFHVDKNG KRLSHIINPN NKRPISHNLA SISVVADSAM
     301  TADGLSTGLF VLGETEALKL AEREKLAVFL IVRDKGYRT AMSSEFEKLL
     351  R*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF111 shows 96.9% identity over a 351aa overlap with an ORF (ORF111a) from strain A of *N.*

meningitidis:

5	orf111a.pep	10 20 30 40 50 60	MPSETRLPNFIRTLIFALSFI FLNACSEQT AQT VTLQGETMGTTYTVKYLSNNRDXLESP
	orf111	10 20 30 40 50 60	MPSETRLPNFIRVLIFALGFIFLNACSEQT AQT VTLQGETMGTTYTVKYLSNNRDKLPSP
10	orf111a.pep	70 80 90 100 110 120	AEIQXRIDDALKEVNRQMSTYQPDSEISR FNQHTAGKPLRISDFAHVTA EAVHLNRLTH
	orf111	70 80 90 100 110 120	AEIQXRIDDALKEVNRQMSTYQPDSEISR FNQHTAGKPLRISDFAHVTA EAVHLNRLTH
15	orf111a.pep	130 140 150 160 170 180	GALDVTVGPLVNLWGFGPDKSVTREPSPEQIKQAASYTGIDKIILKQGKDYASLSKTHPK
	orf111	130 140 150 160 170 180	GALDVTVGPLVNLWGFGPDKSVTREPSPEQIKQAASYTGIDKIILKQGKDYASLSKTHPK
20	orf111a.pep	190 200 210 220 230 240	AYLDLSSIAKGFVDXVAGELEKYGIQNYLVEIGGELHGKXKNARGE PWRIGIEQPNIVQ
	orf111	190 200 210 220 230 240	AYLDLSSIAKGFVDXVAGELEKYGIQNYLVEIGGELHGKXKNARGE PWRIGIEQPNIVQ
30	orf111a.pep	250 260 270 280 290 300	GGNTQIIVPLNNRSXATSGDYRIFHVDKSGKRLSHIINPNNKRPI SHNLASISVXADSAM
	orf111	250 260 270 280 290 300	GGNTQIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNNKRPI SHNLASISVVADSAM
35	orf111a.pep	310 320 330 340 350	TADGXSTGLFVLGETEALKLAEREKLAVFLIVRDKGGYRTAMSSEFEKLLRX
	orf111	310 320 330 340 350	TADGLSTGLFVLGETEALKLAEREKLAVFLIVRDKGGYRTAMSSEFEKLLRX

The complete length ORF111a nucleotide sequence <SEQ ID 447> is:

40	1	ATGCCGTCG	AAACACGCCT	GCCGAACCTT	ATCCGCACCT	TGATATTTGC
	51	CCTGAGTTT	ATCTTCCTGA	ACGCCTGTTC	GGAACAAACC	GCGCAAACCG
	101	TTACCCTGCA	AGGTGAAACG	ATGGGCACGA	CCTATACCGT	CAAATACCTT
	151	TCAAATAATC	GGGACNAACT	CCNCTCACCT	GCCGAAATAC	AAAANCGCAT
	201	CGATGACGCG	CTTAAAGAAG	TCAACCGGCA	GATGTCCACC	TATCAGCCCG
45	251	ACTCCGAAAT	CAGCCGGTTC	AACCAACACA	CAGCCGGCAA	GCCCCTCCGC
	301	ATTTCAAGCG	ACTTCGCACA	CGTTACTGCC	GAAGCCGTCC	ACCTGAACCG
	351	CCTGACACAC	GGCGCGCTGG	ACGTAACCGT	CGGCCCTTG	GTCAACCTTT
	401	GGGGATTCCG	CCCCGACAAA	TCCGTTACCC	GTGAACCGTC	GCCGGAACAA
	451	ATCAAACAAG	CAGCATCTTA	TACGGGCATA	GACAAATCA	TTTGAACA
50	501	AGGCAAAGAT	TACGCTTCCT	TGAGCAAAAC	CCACCCCAAG	GCCTATTTGG
	551	ATTTATCTTC	GATTGCCAAA	GGCTTCGGCG	TTGATNANGT	TGCGGGCGAA
	601	CTGGAAAAAT	ACGGCATTCA	AAATTATCTG	GTCGAAATCG	GCGGNGAGTT
	651	GCACGGCAAA	GNCAAAAACG	CGCGCGGCGA	ACCTTGGCGC	ATCGGCATCG
	701	AACAGCCCAA	CATCGTCCAA	GGCGGCAATA	CGCAGATTAT	CGTCCCGCTG
55	751	AACAACCGTT	CGNTTGCCAC	TTCCGGCGAT	TACCGTATT	TCCACGTCGA
	801	TAAAGCGGC	AAACGCCTCT	CCCATATCAT	TAATCCGAAC	AACAAACGAC
	851	CCATCAGCCA	CAACCTCGCC	TCCATCAGCG	TGNTCGCAGA	CAGTGCGATG
	901	ACGGCGGACG	GCTTNTCCAC	AGGATTATTC	GTATTGGGCG	AAACCGAAGC
	951	CTTAAAGCTG	GCAGAGCGCG	AAAACTCGC	TGTTTTCCTG	ATTGTCAGGG
60	1001	ATAAAGGCGG	CTACCGCACC	GCCATGTCTT	CCGAATTTGA	AAAACGCTC
	1051	CGCTAA				

This encodes a protein having amino acid sequence <SEQ ID 448>:

```

1  MPSETRLPNF IRTLIFALS IFLNACSEQT AQT VTLQGET MGTTYTVKYL
51  SNNRDXLPSP AEIQXRIDDA LKEVNRQMST YQPDSEISRF NQHTAGKPLR

```

101 ISSDFAHVTA EAVHLNRLTH GALDVTVGPL VNLWGFDPK SVTREPSPEQ
 151 IKQAAASYTGI DKIIKQKGD YASLSKTHPK AYLDLSSIAK GFGVDXVAGE
 201 LEKYGIQNYL VEIGGELHGK XKNARGEPWR IGIEQPNIVQ GGNTQIIIVPL
 251 NNRSXATSGD YRIFHVDKSG KRLSHIINPN NKRPISHNLA SISVKADSAM
 301 TADGXSTGLF VLGETEALKL AEREKLAVFL IVRDKGGYRT AMSSEFEKLL
 351 R*

Homology with a predicted ORF from *N.gonorrhoeae*

ORF111 shows 96.6% identity over a 351aa overlap with a predicted ORF (ORF111.ng) from *N.*

10 *gonorrhoeae*:

		10	20	30	40	50	60
	orf111ng	MPSETRLPNLIRALIFALGFIFLNACSEQTAQTVTLQGETMGGTTYTVKYLSNNRDKLPSP					
	orf111	MPSETRLPNFIRVLIFALGFIFLNACSEQTAQTVTLQGETMGGTTYTVKYLSNNRDKLPSP					
15		10	20	30	40	50	60
	orf111	AKIQKRIDDALKEVNRQMSYQTDSEISRFNQHTAGKPLRISSDFAHVTA EAVRLNRLTH					
20	orf111	AEIQKRIDDALKEVNRQMSYQDPSEISRFNQHTAGKPLRISSDFAHVTA EAVRLNRLTH					
		70	80	90	100	110	120
	orf111ng	GALDVTVGPLVNLWGFDPKSVTREPSPEQIKQAASYTGIDKIILOQKDYASLSKTHPK					
25	orf111	GALDVTVGPLVNLWGFDPKSVTREPSPEQIKQAASYTGIDKIILOQKDYASLSKTHPK					
		130	140	150	160	170	180
	orf111ng	AYLDLSSIAKGFVGDKVAGELEKYGIQNYLVEIGGELHGKGNARGEPWRIGIEQPNIIQ					
30	orf111	AYLDLSSIAKGFVGDKVAGELEKYGIQNYLVEIGGELHGKGNARGEPWRIGIEQPNIVQ					
		190	200	210	220	230	240
	orf111ng	GGNTQIIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNKRPISHNLASISVVSDSAM					
35	orf111	GGNTQIIIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNKRPISHNLASISVVADSAM					
		250	260	270	280	290	300
	orf111ng	TADGLSTGLFVLGETEALRLAEQEKLAFLIVRDKDGYRTAMSSEFAKLLRX					
40	orf111	TADGLSTGLFVLGETEALKLAEREKLAVFLIVRDKGGYRTAMSSEFEKLLRX					
		310	320	330	340	350	

The complete length ORF111ng nucleotide sequence <SEQ ID 449> is:

	1	ATGCCGCTCTG	AAACACGCCT	GCCGAACCTT	ATCCGCGCCT	TGATATTTGC
	51	CCTGGGTTTC	ATCTTCCTGA	ACGCCTGTTC	GGaacaacC	GCGCAaacCg
50	101	TTACCCTGCA	AGGCGAAAcg	atGGGTACGA	CCTATACCGT	CAAATACCTT
	151	TCAAATAATC	GGGCAAACT	CCCCTCCCCT	GCCAAAATAC	AAAAGCGCAT
	201	TGATGATGCG	CTTAAAGAAG	TCAACCGGCA	GATGTCCACC	TACCAGACCG
	251	ATTCCGAAAT	CAGCCGGTTC	AACCAACACA	CAGCCGGCAA	GCCCTCCGC
	301	ATTTCAAGCG	ATTTCGCACA	CGTTACCGCC	GAAGCCGTCC	GCCTGAACCG
	351	CCTGACTCAC	GGCGCACTGG	ACGTAACCGT	CGGCCCTTTG	GTCAACCTTT
55	401	GGGGGTTCGG	CCCCGACAAA	TCCGTTACCC	GTGAACCGTC	GCCGGACAA
	451	ATCAAACAGG	CGGCATCTTA	TACGGGCATA	GACAAAATCA	TTTTGCAACA
	501	AGGCAAAGAT	TACGCTTCCT	TGAGCAAAAC	CCACCCCAAA	GCCTATTTGG
	551	ATTTATCTTC	GATTGCCAAA	GGCTTCGGCG	TTGATAAAGT	TGCGGGCGAA
	601	CTGGAAAAAT	ACGGCATTCA	AAATTATCTG	GTGGAAtcg	gcggcGAGTT
60	651	GCACGGCAAA	GGCAAAAATG	CGCACGGCGA	ACCGTGGCGC	ATCGGTATAG
	701	AGCAACCCAA	TATCATCCAA	GgcgGCAata	CGCAGATTat	cgteccgctg
	751	aaCaaccgtt	cgcttGCCAC	TTCCGGCGAT	TaccgtaTTT	tccacgtcgA
	801	TAAAAAcggc	aaacgccttt	cccacaTCAT	CAATCCCaAC	aacAAACgac
	851	ccATCAGcca	caacctcgcc	tccatcagcg	tggtctcAGA	CAGTGCATG
65	901	ACGGCGGACG	GTTtatCCAC	AGGATTATTT	GTTTtagCGC	AAACCGAAGC
	951	CTTAAGGCTG	GCAGAACAAAG	AAAAACTCGC	TGTTTCTCTA	ATTGTCCGGG

1001 ATAAGGACGG CTACCGCACC GCCATGTCTT CCGAATTTGC CAAGCTGCTC
1051 CGCTAA

This encodes a protein having amino acid sequence <SEQ ID 450>:

5 1 MPSETRLPNL IRALIFALGF IFLNACSEQT AQTVTLOGET MGTITYVKYL
 51 SNNRDKLPSP AKIQKRIDDA LKEVNRQMST YQTDSEISRF NQHTAGKPLR
 101 ISSDFAHVTA EAVRLNRLTH GALDVTVGPL VNLWGFDPK SVTREPSPEQ
 151 IKQAASYTGI DKIILOQGD YASLSKTHPK AYLDLSSIAK GFGVDK VAGE
 201 LEKYGIQNYL VEIGGELHGK GKNAHGEPWR IGIEQPNIIQ GGNTQIIIVPL
 251 NNRSLATSGD YRIFHVDKNG KRLSHIINPN NKRPISHNLA SISVSDSAM
10 301 TADGLSTGLF VLGETEALRL AEQEKLAFL IVRDKDGYRT AMSSEFAKLL
 351 R*

This protein shows homology with a hypothetical lipoprotein precursor from *H. influenzae*:

sp|P44550|YOJL_HAEIN HYPOTHETICAL LIPOPROTEIN HI0172 PRECURSOR >gi|1074292|pir|4
hypothetical protein HI0172 - Haemophilus influenzae (strain Rd KW20)
15 >gi|1573128 (U32702) hypothetical [Haemophilus influenzae] Length = 346
Score = 353 bits (896), Expect = 9e-97
Identities = 181/344 (52%), Positives = 247/344 (71%), Gaps = 4/344 (1%)

Query: 7 LPNLIRALIFALGFIFLNACSEQTAQTVTLOGETMGTITYVKYLSNNRDKLPSPAKIQKR 66
+ LI +I + L AC ++T + ++L G+TMGTTY VKYL + S K +
20 Sbjct: 1 MKKLISGIIAVAMALSLAACQKET-KVISLSGKTMGTTYHVKYLDGSGITATSE-KTHEE 58

Query: 67 IDDALKEVNRQMSTYQTDSEISRFNQHT-AGKPLRISSDFAHVTAEAVRLNRLTHGALDV 125
I+ LK+VN +MSTY+ DSE+SRFNQ+T P+ IS+DFA V AEA+RLN++T GALDV
25 Sbjct: 59 IEAILKDVNAKMSTYKKDSELSRFNQNTQVNTPIEISADFAKVLAEAIRLNKVTGALDV 118

Query: 126 TVGPVLNWLWGFDPKSVTREPSPEQIKQAASYTGIDKIILOQGDYASLSKTHPKAYLDL 185
TVGP+VNLWGFDP+K ++P+PEQ+ + ++ GIDKI L K+ A+LSK P+ Y+DL
30 Sbjct: 119 TVGPVNLWGFDPKPEKQPTPEQLAERQAWVGIDKITLDTNKEKATLSKALPQVYVDL 178

Query: 186 SSIAKGFVGDQVAGELEKYGIQNYLVEIGGELHGKGKNAHGEPWRIGIEQPNIIQGGNTQ 245
SSIAKGFVD+VA +LE+ QNY+VEIGGE+ KGKN G+PW+I IE+P +

Sbjct: 179 SSIAKGFVGDQVAKLEQLNAQNYMVEIGGEIRAKGKNIEGKPWQIAIEKPTTTGERAVE 238
35 Query: 246 IIVPLNNRSLATSGDYRIFHVDKNGKRLSHIINPNKRPISHNLAISVSDSAMTADGL 305
++ LNN +A+SGDYRI+ ++NGKR +H I+P PI H+LASI+V++ ++MTADGL
Sbjct: 239 AVIGLNNMGMASSGDYRIY-FEENGKRAFEIDPKTGYPHQHHLASITVLAPTSMTADGL 297

Query: 306 STGLFVLGETEALRLAEQEKLAFLIVRDKDGYRTAMSSEFAKL 349
STGLFVLGE +AL +AE+ LAV+LI+R +G+ T SS F KL
40 Sbjct: 298 STGLFVLGEDKALEVAEKNLAVYLIIRTDNGFVTKSSSAFKKL 341

Based on this analysis, it is predicted that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 54

45 The following partial DNA sequence was identified in *N. meningitidis* <SEQ ID 451>:

51 1 ..CCGTGCCGCC GACAGGGCGA CGACGTGTAT GCGGCGCAGC CGTCCCCTCA
 51 AAAAATTGTGG CTGCGCTTCA TCGGCGGCCG GTCGCATCAA AATATACGGG
 101 GCGGCGCGGC TCGGACGGG TGGCGCAAAG GCGTCAAAT CGCGCGCGAG
 151 GTGTTTGTAC GGCAAAATGA AGGCAGCCKA yTGCAATCG GCGTGATGGG
50 201 CGGCAGGCC GGCAGCAGC CwTCAGTCAA CGGCAAAGGC GGTGCGGCAG
 251 gCAGTGATTT GTATGGTTAT GgCGGGGgTG TTTATGCTgC GTGGCATCAG
 301 TTGCGCGATA AACAAACGGG TgCGTATTTG GACGGCTGGT TGCAATACCA
 351 ACGTTTCAA CACCGCATCA ATGATGAAAA CCGTGCGGAA CgCTACAAAA
 401 CCAAAGGTTG GACGGCTTCT GTCGAAGGCG GCTACAACGC GCTTGTGGCG
55 451 GAAGGCATTG TCGGAAAGG CAATAATGTG CGGTTTACC TACAACCGCA
 501 GgCGCAGTTT ACCTACTTGG GCGTAAACGG CGGCTTACC GACAGCGAGG
 551 GGACGGCGGT CGGACTGCTC GGCAGCGGTC AGTGGCAAAG CCGCGCCGGC
 601 AtTCGGGCAA AAACCGGTTT TGCTTTGCGT AACGGTGTC ATCTTCAGCC
 651 TTTTGCCGCT TTAATGtTt TGCACAGGTC AAAATCTTTC GGCGTGGA
60 701 TGGACGGCGA AAAACAGACG CTGGCAGGCA GGACGGCACT CGAAGGGCGG

751 TTCGGTATTG AAGCCGGTTG GAAAGGCCAT ATGTCCGCA..

This corresponds to the amino acid sequence <SEQ ID 452; ORF35>:

1 ..PCRRQGDDVY AAHASRQKLW LRFIGGRSHQ NIRGGAADG WRKGVQIGGE
51 VFVRONEGSX LAIGVMGGRA GHASVNGKG GAAGSDLYGY GGGVYAAWHQ
101 LRDQKTGAYL DGWLQYQRFK HRINDENRAE RYKTKGTAS VEGGYNALVA
151 EGIVGKGNNV RFYLQPPAQF TYLGVNGGFT DSEGTAVGLL GSGQWQSRAG
201 IRAKTRFALR NGVNLQPFPA FNVLHRSKSF GVEMDGEKQT LAGRTALEGR
251 FGIEAGWKGH MSA..

Computer analysis of this amino acid sequence gave the following results:

10 Homology with putative secreted VirG-homologue of *N. meningitidis* (accession number A32247)

ORF and virg-h protein show 51% aa identity in 261aa overlap:

Orf35 5 QGDDVYAAHASRQKLWLRFIGGRSHQNIIRGGA-ADGWRKGVQIGGEVFVRONEGSXLAI 63
+ D++ R+ LWLR I G S+Q ++G A +G+RKGVQ+GGEVF QNE + L+I
virg-h 396 KNSDIFDRTLPRKGLWLRVIDGHSNQWVQKGTAPVEGYRKGVQLGGEVFTWQNESNQLSI 455
15 Orf35 64 GVMGGRAGQHASVNGKG--GAAGSDLYGYGGGVYAAWHQLRDQKTGAYLDGWLQYQRFKH 121
G+MGG+A Q ++ + ++ G+G GVYA WHQL+DKQTGAY D W+QYQRF+H
virg-h 456 GLMGGQAEQRSTFHNPDNDLTTGNVKGFGAGVYATWHQLQDKQTGAYADSWMQYQRFH 515
20 Orf35 122 RINDENRAERYKTKGTASVEGGYNALVAEGIVGKGNNVRFYLQPPAQFTYLGVNGGFTD 181
RIN E+ ER+ +KG TAS+E GYNAL+AE KGN++R YLQPPAQ TYLGVNG F+D
virg-h 516 RINTEDGTERFTSKGITASIEAGYNALLAEHFTKKGNLSRVYLQPPAQLTYLGVNGKFSD 575
25 Orf35 182 SEGTAVGLLGGSGQWQSRAGIRAKTRFALRNGVNLQPFPAFNVLHRSKSFGEVEMDGEKQTL 241
SE V LLGS Q Q+R G++AK +F+L + ++PFAA N L+ +K FGVEMDGE++ +
virg-h 576 SENAHVNLLGSRQLQTRVGVQAKAQFSLYKNIAIEPFAAVNALYHNKPFGEVEMDGERRVI 635
Orf35 242 AGRTALEGRFGIEAGWKGHMS 262
+TA+E + G+ K H++
30 virg-h 636 NNKTAIESQLGVAVKIKSHLT 656

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF35 shows 96.9% identity over a 259aa overlap with an ORF (ORF35a) from strain A of *N.*

meningitidis:

35 orf35.pep 10 20 30
PCRRQGDDVYAAHASRQKLWLRFIGGRSHQNIIRG
orf35a QRLAIPAEAVLYAQQAYAANTLFLGLRAADRGDDVYAADPSRQKLWLRFIGGRSHQNIIRG
310 320 330 340 350 360
40 orf35.pep 40 50 60 70 80 90
GAAADGWRKGVQIGGEVFVRONEGSXLAI GVMGGRAGQHASVNGKGGAAGSDLYGYGGGV
orf35a GAAADGRRKGVQIGGEVFVRONEGSRLAIGVMGGRAGQHASVNGKGGAAGSYLHGYGGGV
370 380 390 400 410 420
45 orf35.pep 100 110 120 130 140 150
YAAWHQLRDQKTGAYLDGWLQYQRFKHRINDENRAERYKTKGTASVEGGYNALVAEGIV
orf35a YAAWHQLRDQKTGAYLDGWLQYQRFKHRINDENRAERYKTKGTASVEGGYNALVAEGVV
430 440 450 460 470 480
50 orf35.pep 160 170 180 190 200 210
GKGNNVRFYLQPPAQFTYLGVNGGFTDSEGTAVGLLGGSGQWQSRAGIRAKTRFALRNGVN
orf35a GKGNNVRFYLQPPAQFTYLGVNGGFTDSEGTAVGLLGGSGQWQSRAGIRAKTRFALRNGVN
490 500 510 520 530 540
55 orf35.pep 220 230 240 250 260
LQPFPAFNVLHRSKSFGEVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSA
orf35a LQPFPAFNVLHRSKSFGEVEMDGEKQTLAGRTALEGRFGIEAGWKGHMSA
60

-272-

orf35a LQPF AAFNVLHRSKSPGVEMDGEKQTLA GR TALEGRFGIEAGWKGHMSARIGYGKRTDGD
550 560 570 580 590 600

5 orf35a KEAALS LKWLFX
610 620

The complete length ORF35a nucleotide sequence <SEQ ID 453> is:

```

1  ATGTT CAGAG CTCAGCTTGG TTCAAATACT CGTTCTACCA AAATCGGCGA
51  CGATGCCGAT TTTTCATTTT CAGACAAGCC GAAACCCGGC ACTTCCCATT
101 ATTTTTC CAG CGGTAAAACC GATCAAAATT CATCCGAATA TGGGTATGAC
151 GAAATCAATA TCCAAGGTAA AAATAACAAT AGCGGCATAC TCGCCGTCGA
201 TAATATGCCC GTTGTTAAGA AATATATTAC AGATACTTAC GGGGATAATT
251 TAAAGGATGC GGTTAAGAAG CAATTACAGG ATTTATACAA AACAAAGACCC
301 GAAGCTTGGG AAGAAAATAA AAAACGGACT GAGGAGGCGT ATATAGAACA
351 GCTTGGACCA AAATTTAGTA TACTCAACA GAAAAACCCG GATTTAATTA
15 401 ATAAATTGGT AGAAGATTCC GTACTCACTC CTCATAGTAA TACATCACAG
451 ACTAGTCTCA ACAACATCTT CAATAAAAAA TTACACGTCA AAATCGAAAA
501 CAAATCCCAC GTCGCCGGAC AGGTGTTGGA ACTGACCAAG ATGACGCTGA
551 AAGATTCCCT TTGGGAACCG CGCCGCCATT CCGACATCCA TATGCTGGAA
601 ACTTCCGATA ATGCCCGCAT CCGCTGAAC ACGAAAGATG AAAAAGTAC
20 651 CGTCCATAAA GCGTATCAGG GCGGTGCGGA TTTCTGTTC GGCTACGACG
701 TGCGGGAGTC GGACAAACCC GCCCTGACCT TTGAAGAAAA AGTCAGCGGA
751 CAATCCGGCG TGGTTTGGGA ACGCCGGCCG GAAAATCTGA AAACGCTCGA
801 CGGGCGCAAA CTGATTGCGG CGGAAAAGGC AGACTCTAAT TCGTTGCGT
851 TTAACAAAAA TTACCGGCAG GACTGTACG AATTATTGCT CAAGCAATGC
25 901 GAAGGCGGAT TTTGCTTGGG CGTGCAGCGT TTGGCTATCC CCGAGGCGGA
951 AGCGGTTTTA TATGCCCAAC AGGCTTATGC GGCAATACT TTGTTCCGGC
1001 TGCGTGCCGC CGACAGGGGC GACGACGTGT ATGCCGCCGA TCCGTCCCGT
1051 CAAAAATTGT GGCTGCGCTT CATCGGCGGC CGGTGCGATC AAAATATACG
1101 GGGCGGCGCG GCTGCGGACG GCGGCGCAA AGGCGTGCAA ATCGGCGGCG
30 1151 AGGTGTTTGT ACGGCAAAAT GAAGGCAGCC GGCTGGCAAT CGGCGTGATG
1201 GCGGCGAGGG CTGGCCAGCA CGCATCAGT AACGGCAAAG GCGGTGCGGC
1251 AGGCAGTTAT TTGCATGGTT ATGGCGGGGG TGTTTATGCT GCGTGGCATC
1301 AGTTGCGCGA TAAACAAACG GGTGCGTATT TGGACGGCTG GTTGCAATAC
1351 CAACGTTTCA AACACCGCAT CAATGATGAA AACCGTGCGG AACGCTACAA
35 1401 AACCAAAGGT TGGACGGCTT CTGTGAAGG CCGCTACAAC GCGCTGTGG
1451 CGGAAGGCGT TGTGCGAAAA GGCAATAATG TGCGGTTTTA CCTGCAACCG
1501 CAGGCGCAGT TTACCTACTT GGGCGTAAAC GCGGCTTTA CCGACAGCGA
1551 GGGGACGGCG GTCGGACTGC TCGGACGGG TCACTGGCAA AGCCGCGCCG
40 1601 GCATTGCGGC AAAAACCCGT TTTGCTTTGC GTACGGTGT CAATCTTCAG
1651 CCTTTGCGC CTTTAAATGT TTTGCACAGG TCAAAATCTT TCGGCGTGGA
1701 AATGGACGGC GAAAAACAGA CGCTGGCAGG CAGGACGGCG CTCGAAGGGC
1751 GGTTCGGCAT TGAAGCCGGT TGGAAAGGCC ATATGTCCGC ACGCATCGGA
1801 TACGGCAAAA GGACGGACGG CGACAAAGAA GCCGCATTGT CGCTCAAATG
1851 GCTGTTTTGA

```

45 This encodes a protein having amino acid sequence <SEQ ID 454>:

```

1  MFRAQLGSNT RSTKIGDDAD FSFSKPKPG TSHYFSSGKT DQNSSEYGYD
51  EINIQGNYN SGILAVDNMP VVKYITDITY GDNLDKAVKK QLQDLYKTRP
101  EAWREENKRT EEAYIEQLGP KFSILKQKNP DLINKLVEDS VLTPHSNTSQ
50 151  TSLNNIFNKK LHVKIENKSH VAGQVLELTK MTLKDSLWEP RRHSDIHMLE
201  TSDNARIRLN TKDEKLTVHK AYQGGADFLF GYDVRESDKP ALTFEEKVSG
251  QSGVVLERRP ENLKTLDGRK LIAAEKADSN SFAFKQNYRQ GLYELLKQC
301  EGGFCLGVQR LAIPEAEAVL YAQQAYAANT LFGLRAADRG DDVYAADPSR
351  OKLWLRFIGG RSHQNIIRGA AADGRKGVQ IGGEVFVRQN EGSRLAIGVM
401  GGRAGQHASV NGKGAAGSY LHGYGGGVYA AWHQLRDKQT GAYLDGWLQY
55 451  QRFKHRINDE NRAERYTKG WTASVEGGYN ALVAEGVVGK GNNVRFYLPQ
501  QAQFTYLGVN GGFTDSEGT VLLSGSQWQ SRAGIRAKTR FALRNGVNLQ
551  PFAAFNVLHR SKSFGVEMDG EKQTLA GR TALEGRFGIEAG WKGHMSARIG
601  YGKRTDGDKE AALS LKWLFX*

```

Homology with a predicted ORF from *N.gonorrhoeae*

60 ORF35 shows 51.7% identity over a 261aa overlap with a predicted ORF (ORF35ngh) from *N.*

gonorrhoeae:

```

orf35.pep PCRRQDDVYAAHASRQKLWLRFIGGRSHQNIIRG 34
:::|:: |::|::|::|::|::|
orf35ngh FTKVQERDDIAIYAQQAQAANTLFALRLNDKNSDIFDRTLPRKGLWLRVIDGHSNQWVQG 370

```

	orf35.pep	GAA-ADGWRKGVQIGGEVFRQNEGSXLAIGVMGGRAGQHASVNGKG--GAAGSDLYGYG	91
	orf35ngh	KTAPVEGYRKGVLGGEVFTWQNESNQLSIGLMGGQAEQRSTFRNPDTDNLTGNVKGFG	430
5	orf35.pep	GGVYAAWHQLRDKQTGAYLDGWLQYQRFKHRINDENRAERYKTKGWTASVEGGYNALVAE	151
	orf35ngh	AGVYATWHQLQDKQTGAYVDSWMQYQRFHRINTEYATERFTSKGITASIEAGYNALLAE	490
10	orf35.pep	GIVGKGNVRFYLPQQAQFTYLVNGGFTDSEGTAVGLLGSGQWQSRAGIRAKTRFALRN	211
	orf35ngh	HFTKKGNSLRVYLQPOAQLTYLVNGKFSSENAQVNLLGSRQLQSRVGVQAKAQFAFTN	550
15	orf35.pep	GVNLQPFPAFNVLHRSKSGFVEMDGEKOTLAGRTALEGRFGIEAGWKGHMSA	263
	orf35ngh	GVTQPFVAVNSIYQQKPFGEIDGRRRVINNKTVIETQLGVAARIKSHLTQASFNRT	610

A partial ORF35ngh nucleotide sequence <SEQ ID 455> is predicted to encode a protein having partial amino acid sequence <SEQ ID 456>:

	1	..KKLRDRNSEY	WKEETYHIKS	NGRTYPNIPA	LFPKHFPDPF	ENINNSKKIS
20	51	FYDKEYTEDY	LVGFARGFGV	EKRNGEEK	LRQYFKDCVN	TENSNDNCK
	101	ISSFGNYGPI	LIKSDIFALA	SQIKNSHINS	EILSVGNIE	WLRPTLNKLT
	151	GWQEHLYAGL	DPFHYIEVTD	NSHVIGQID	LGALELTNSL	WKPRWNSNID
	201	YLITKNAEIR	FNTKNESLLV	KEDYAGGARF	RFAYDLKDKV	PEIPVLTFEK
	251	NITGTSDIIF	EGKALDNLKH	LDGHQIVKVN	DTADKDAFRL	SSKYRKGIYT
25	301	LSLQQRPEGF	FTKVQERDDI	AIYAQQAQAA	NTLFALRLND	KNSDIFDRTL
	351	PRKGLWLRVI	DGHSNQWVQG	KTAPVEGYRK	GVQLGGEVFT	WQNESNQLSI
	401	GLMGGQAEQR	STFRNPDTDN	LTTGNVKGFG	AGVYATWHQL	QDKQTGAYVD
	451	SWMQYQRFHR	RINTEYATER	FTSKGITASI	EAGYNALLAE	HFTKKGNSLR
	501	VYLQPOAQLT	YLVNGKFS	SENAQVNLLG	SRQLQSRVGV	QAKAQFAFTN
30	551	GVTQPFVAV	NSIYQQKPF	VEIDGRRVI	NNKTVIETQL	GVAARIKSHL
	601	TLQASFNRT	SKHHAKQGA	LNLQWTF*		

Based on this prediction, these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 55

35 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 457>:

	1	..GCGGAATATG	TTCAGTTCTC	TATAGATTTG	TTCAGTGTGG	GTAATCGGG
	51	GGGCGGTATA	CCTAAGGCTA	AGCCTGTGTT	TGATGCGAAA	CCGAGATGGG
	101	AGGTTGATAG	GAAGCTTAAT	AAATGACAA	CTCGTGAGCA	GGTGAGAGAA
40	151	AATGTTTCAGG	AAACGAGAAG	AAGGAGTCAG	AGTAGTCAGT	TTAAAGCCCCA
	201	TGCGCAACGA	GAATGGGAAA	ATAAAACAGG	GTTAGATTTT	AATCATTTTA
	251	TAGGTGGTGA	TATCAATAAA	AAAGGCACAG	TAACAGGAGG	GCATAGTCTA
	301	ACCCGTGGTG	ATGTACGGGT	GATACAACAA	ACCTCGGCAC	CTGATAAACA
	351	TGGGGT.TTA	TCAAGCGACA	GTGGAAATTN	A	

This corresponds to the amino acid sequence <SEQ ID 458; ORF46>:

45	1	..AEYVQFSIDL	FSVGKSGGGI	PKAKPVFDAK	PRWEVDRKLN	KLTTRQVEK
	51	NVQETRRRSQ	SSQFKAHAQR	EWENKTGLDF	NHFIGGDINK	KGTVTGGHSL
	101	TRGDVRVIOQ	TSAPDKHXL	SSDSGNX		

Further work revealed further partial nucleotide sequence <SEQ ID 459>:

50	1	..GCAGTGTGCC	TnCCGATGCA	TGCACACGCC	TCAaATTGG	CAAACGATT
	51	TTTTATCCGG	CAGGTTCTCG	ACCGTCAGCA	TTTCGAACCC	GACGGGAAAT
	101	ACCACCTATT	CGGCAGCAGG	GGGGAACCTG	CCGAGCGCCA	GTCTCATATC
	151	GGATTGGGAA	AAATACAAAG	CCATCAGTTG	GGCAACCTGA	TGATTCAACA
	201	GGCGGCCATT	AAAGGAAATA	TCCGCTACAT	TGTCCGCTTT	TCCGATCAGG
	251	GGCACGAAGT	CCATTCCCCs	TTCGACAACC	ATGCCTCACA	TCCGATTCT
55	301	GATGAAGCCG	GATGTCCTG	TGACGGATT	AGCCTTTACC	GCATCCATTG
	351	GGACGGATAC	GAACACCATC	CCGCCGACGG	CTATGACGGG	CCACAGGGCG
	401	GCGGTATCC	CGCTCCCAA	GGCGCGAGGG	ATATATACAG	TTACGACATA

-274-

5
10
15
20
25
30
35
40
45
50
55
60
65

451 AAGGCGTTG CCCAAATAT CCGCCTCAAC CTGACCGACA ACCGCAGCAC
501 CGGACAACGG CTTGCCGACC GTTCCACAA TGCCGGTAGT ATGCTGACGC
551 AAGGAGTAGG CGACGGATTC AAACGCGCCA CCCGATACAG CCCCAGCTG
601 GACAGATCGG GCAATGCCGC CGAAGCCTTC AACGGCACTG CAGATATCGT
651 TAAAAACATC ATCGGCGCTG CAGGAGAAAT TGT

This corresponds to the amino acid sequence <SEQ ID 460; ORF46-1>:

10
15
20
25
30
35
40
45
50
55
60
65

1 ..AVCLPMHAHA SXLANDSFIR QVLDROHFEP DGKYHLFGSR GELAERQSHI
51 GLGKIQSHQL GNLMIQAAAI KGNIGYIVRF SDHGHEVHSP FDNHASHSDS
101 DEAGSPVDGF SLYRIHWDGY EHHPADGYDG PGGGGYPAPK GARDIYSYDI
151 KGVAQNIRLN LTDNRSTGQR LADRFHNAGS MLTQGVGDGF KRATRYSPLE
201 DRSGNAAEAF NGTADIVKNI IGAGEI

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF46 shows 98.2% identity over a 111aa overlap with a predicted ORF (ORF46ng) from *N.*

15 *gonorrhoeae*:

20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000
1005
1010
1015
1020
1025
1030
1035
1040
1045
1050
1055
1060
1065
1070
1075
1080
1085
1090
1095
1100
1105
1110
1115
1120
1125
1130
1135
1140
1145
1150
1155
1160
1165
1170
1175
1180
1185
1190
1195
1200
1205
1210
1215
1220
1225
1230
1235
1240
1245
1250
1255
1260
1265
1270
1275
1280
1285
1290
1295
1300
1305
1310
1315
1320
1325
1330
1335
1340
1345
1350
1355
1360
1365
1370
1375
1380
1385
1390
1395
1400
1405
1410
1415
1420
1425
1430
1435
1440
1445
1450
1455
1460
1465
1470
1475
1480
1485
1490
1495
1500
1505
1510
1515
1520
1525
1530
1535
1540
1545
1550
1555
1560
1565
1570
1575
1580
1585
1590
1595
1600
1605
1610
1615
1620
1625
1630
1635
1640
1645
1650
1655
1660
1665
1670
1675
1680
1685
1690
1695
1700
1705
1710
1715
1720
1725
1730
1735
1740
1745
1750
1755
1760
1765
1770
1775
1780
1785
1790
1795
1800
1805
1810
1815
1820
1825
1830
1835
1840
1845
1850
1855
1860
1865
1870
1875
1880
1885
1890
1895
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040
2045
2050
2055
2060
2065
2070
2075
2080
2085
2090
2095
2100
2105
2110
2115
2120
2125
2130
2135
2140
2145
2150
2155
2160
2165
2170
2175
2180
2185
2190
2195
2200
2205
2210
2215
2220
2225
2230
2235
2240
2245
2250
2255
2260
2265
2270
2275
2280
2285
2290
2295
2300
2305
2310
2315
2320
2325
2330
2335
2340
2345
2350
2355
2360
2365
2370
2375
2380
2385
2390
2395
2400
2405
2410
2415
2420
2425
2430
2435
2440
2445
2450
2455
2460
2465
2470
2475
2480
2485
2490
2495
2500
2505
2510
2515
2520
2525
2530
2535
2540
2545
2550
2555
2560
2565
2570
2575
2580
2585
2590
2595
2600
2605
2610
2615
2620
2625
2630
2635
2640
2645
2650
2655
2660
2665
2670
2675
2680
2685
2690
2695
2700
2705
2710
2715
2720
2725
2730
2735
2740
2745
2750
2755
2760
2765
2770
2775
2780
2785
2790
2795
2800
2805
2810
2815
2820
2825
2830
2835
2840
2845
2850
2855
2860
2865
2870
2875
2880
2885
2890
2895
2900
2905
2910
2915
2920
2925
2930
2935
2940
2945
2950
2955
2960
2965
2970
2975
2980
2985
2990
2995
3000
3005
3010
3015
3020
3025
3030
3035
3040
3045
3050
3055
3060
3065
3070
3075
3080
3085
3090
3095
3100
3105
3110
3115
3120
3125
3130
3135
3140
3145
3150
3155
3160
3165
3170
3175
3180
3185
3190
3195
3200
3205
3210
3215
3220
3225
3230
3235
3240
3245
3250
3255
3260
3265
3270
3275
3280
3285
3290
3295
3300
3305
3310
3315
3320
3325
3330
3335
3340
3345
3350
3355
3360
3365
3370
3375
3380
3385
3390
3395
3400
3405
3410
3415
3420
3425
3430
3435
3440
3445
3450
3455
3460
3465
3470
3475
3480
3485
3490
3495
3500
3505
3510
3515
3520
3525
3530
3535
3540
3545
3550
3555
3560
3565
3570
3575
3580
3585
3590
3595
3600
3605
3610
3615
3620
3625
3630
3635
3640
3645
3650
3655
3660
3665
3670
3675
3680
3685
3690
3695
3700
3705
3710
3715
3720
3725
3730
3735
3740
3745
3750
3755
3760
3765
3770
3775
3780
3785
3790
3795
3800
3805
3810
3815
3820
3825
3830
3835
3840
3845
3850
3855
3860
3865
3870
3875
3880
3885
3890
3895
3900
3905
3910
3915
3920
3925
3930
3935
3940
3945
3950
3955
3960
3965
3970
3975
3980
3985
3990
3995
4000
4005
4010
4015
4020
4025
4030
4035
4040
4045
4050
4055
4060
4065
4070
4075
4080
4085
4090
4095
4100
4105
4110
4115
4120
4125
4130
4135
4140
4145
4150
4155
4160
4165
4170
4175
4180
4185
4190
4195
4200
4205
4210
4215
4220
4225
4230
4235
4240
4245
4250
4255
4260
4265
4270
4275
4280
4285
4290
4295
4300
4305
4310
4315
4320
4325
4330
4335
4340
4345
4350
4355
4360
4365
4370
4375
4380
4385
4390
4395
4400
4405
4410
4415
4420
4425
4430
4435
4440
4445
4450
4455
4460
4465
4470
4475
4480
4485
4490
4495
4500
4505
4510
4515
4520
4525
4530
4535
4540
4545
4550
4555
4560
4565
4570
4575
4580
4585
4590
4595
4600
4605
4610
4615
4620
4625
4630
4635
4640
4645
4650
4655
4660
4665
4670
4675
4680
4685
4690
4695
4700
4705
4710
4715
4720
4725
4730
4735
4740
4745
4750
4755
4760
4765
4770
4775
4780
4785
4790
4795
4800
4805
4810
4815
4820
4825
4830
4835
4840
4845
4850
4855
4860
4865
4870
4875
4880
4885
4890
4895
4900
4905
4910
4915
4920
4925
4930
4935
4940
4945
4950
4955
4960
4965
4970
4975
4980
4985
4990
4995
5000
5005
5010
5015
5020
5025
5030
5035
5040
5045
5050
5055
5060
5065
5070
5075
5080
5085
5090
5095
5100
5105
5110
5115
5120
5125
5130
5135
5140
5145
5150
5155
5160
5165
5170
5175
5180
5185
5190
5195
5200
5205
5210
5215
5220
5225
5230
5235
5240
5245
5250
5255
5260
5265
5270
5275
5280
5285
5290
5295
5300
5305
5310
5315
5320
5325
5330
5335
5340
5345
5350
5355
5360
5365
5370
5375
5380
5385
5390
5395
5400
5405
5410
5415
5420
5425
5430
5435
5440
5445
5450
5455
5460
5465
5470
5475
5480
5485
5490
5495
5500
5505
5510
5515
5520
5525
5530
5535
5540
5545
5550
5555
5560
5565
5570
5575
5580
5585
5590
5595
5600
5605
5610
5615
5620
5625
5630
5635
5640
5645
5650
5655
5660
5665
5670
5675
5680
5685
5690
5695
5700
5705
5710
5715
5720
5725
5730
5735
5740
5745
5750
5755
5760
5765
5770
5775
5780
5785
5790
5795
5800
5805
5810
5815
5820
5825
5830
5835
5840
5845
5850
5855
5860
5865
5870
5875
5880
5885
5890
5895
5900
5905
5910
5915
5920
5925
5930
5935
5940
5945
5950
5955
5960
5965
5970
5975
5980
5985
5990
5995
6000
6005
6010
6015
6020
6025
6030
6035
6040
6045
6050
6055
6060
6065
6070
6075
6080
6085
6090
6095
6100
6105
6110
6115
6120
6125
6130
6135
6140
6145
6150
6155
6160
6165
6170
6175
6180
6185
6190
6195
6200
6205
6210
6215
6220
6225
6230
6235
6240
6245
6250
6255
6260
6265
6270
6275
6280
6285
6290
6295
6300
6305
6310
6315
6320
6325
6330
6335
6340
6345
6350
6355
6360
6365
6370
6375
6380
6385
6390
6395
6400
6405
6410
6415
6420
6425
6430
6435
6440
6445
6450
6455
6460
6465
6470
6475
6480
6485
6490
6495
6500
6505
6510
6515
6520
6525
6530
6535
6540
6545
6550
6555
6560
6565
6570
6575
6580
6585
6590
6595
6600
6605
6610
6615
6620
6625
6630
6635
6640
6645
6650
6655
6660
6665
6670
6675
6680
6685
6690
6695
6700
6705
6710
6715
6720
6725
6730
6735
6740
6745
6750
6755
6760
6765
6770
6775
6780
6785
6790
6795
6800
6805
6810
6815
6820
6825
6830
6835
6840
6845
6850
6855
6860
6865
6870
6875
6880
6885
6890
6895
6900
6905
6910
6915
6920
6925
6930
6935
6940
6945
6950
6955
6960
6965
6970
6975
6980
6985
6990
6995
7000
7005
7010
7015
7020
7025
7030
7035
7040
7045
7050
7055
7060
7065
7070
7075
7080
7085
7090
7095
7100
7105
7110
7115
7120
7125
7130
7135
7140
7145
7150
7155
7160
7165
7170
7175
7180
7185
7190
7195
7200
7205
7210
7215
7220
7225
7230
7235
7240
7245
7250
7255
7260
7265
7270
7275
7280
7285
7290
7295
7300
7305
7310
7315
7320
7325
7330
7335
7340
7345
7350
7355
7360
7365
7370
7375
7380
7385
7390
7395
7400
7405
7410
7415
7420
7425
7430
7435
7440
7445
7450
7455
7460
7465
7470
7475
7480
7485
7490
7495
7500
7505
7510
7515
7520
7525
7530
7535
7540
7545
7550
7555
7560
7565
7570
7575
7580
7585
7590
7595
7600
7605
7610
7615
7620
7625
7630
7635
7640
7645
7650
7655
7660
7665
7670
7675
7680
7685
7690
7695
7700
7705
7710
7715
7720
7725
7730
7735
7740
7745
7750
7755
7760
7765
7770
7775
7780
7785
7790
7795
7800
7805
7810
7815
7820
7825
7830
7835
7840
7845
7850
7855
7860
7865
7870
7875
7880
7885
7890
7895
7900
7905
7910
7915
7920
7925
7930
7935
7940
7945
7950
7955
7960
7965
7970
7975
7980
7985
7990
7995
8000
8005
8010
8015
8020
8025
8030
8035
8040
8045
8050
8055
8060
8065
8070
8075
8080
8085
8090
8095
8100
8105
8110
8115
8120
8125
8130
8135
8140
8145
8150
8155
8160
8165
8170
8175
8180
8185
8190
8195
8200
8205
8210
8215
8220
8225
8230
8235
8240
8245
8250
8255
8260
8265
8270
8275
8280
8285
8290
8295
8300
8305
8310
8315
8320
8325
8330
8335
8340
8345
8350
8355
8360
8365
8370
8375
8380
8385
8390
8395
8400
8405
8410
8415
8420
8425
8430
8435
8440
8445
8450
8455
8460
8465
8470
8475
8480
8485
8490
8495
8500
8505
8510
8515
8520
8525
8530
8535
8540
8545
8550
8555
8560
8565
8570
8575
8580
8585
8590
8595
8600
8605
8610
8615
8620
8625
8630
8635
8640
8645
8650
8655
8660
8665
8670
8675
8680
8685
8690
8695
8700
8705
8710
8715
8720
8725
8730
8735
8740
8745
8750
8755
8760
8765
8770
8775
8780
8785
8790
8795
8800
8805
8810
8815
8820
8825
8830
8835
8840
8845
8850
8855
8860
8865
8870
8875
8880
8885
8890
8895
8900
8905
8910
8915
8920
8925
8930
8935
8940
8945
8950
8955
8960
8965
8970
8975
8980
8985
8990
8995
9000
9005
9010
9015
9020
9025
9030
9035
9040
9045
9050
9055
9060
9065
9070
9075
9080
9085
9090
9095
9100
9105
9110
9115
9120
9125
9130
9135
9140
9145
9150
9155
9160
9165
9170
9175
9180
9185
9190
9195
9200
9205
9210
9215
9220
9225
9230
9235
9240
9245
9250
9255
9260
9265
9270
9275
9280
9285
9290
9295
9300
9305
9310
9315
9320
9325
9330
9335
9340
9345
9350
9355
9360
9365
9370
9375
9380
9385
9390
9395
9400
9405
9410
9415
9420
9425
9430
9435
9440
9445
9450
9455
9460
9465
9470
9475
9480
9485
9490
9495
9500
9505
9510
9515
9520
9525
9530
9535
9540
9545
9550
9555
9560
9565
9570
9575
9580
9585
9590
9595
9600
9605
9610
9615
9620
9625
9630
9635
9640
9645
9650
9655
9660
9665
9670
9675
9680
9685
9690
9695
9700
9705
9710
9715
9720
9725
9730
9735
9740
9745
9750
9755

-275-

1151 GTTACGGCAA AGAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGC
 1201 AAAAATGTCA AACTGGCAGA CCAACGCCAC CCGAAGACAG GCGTACCGTT
 1251 TGACGGTAAA GGGTTTCCGA ATTTTGAGAA GCACGTGAAA TATGATACGA
 1301 AGCTCGATAT TCAAGAATTA TCGGGGGGCG GTATACCTAA GGCTAAGCCT
 1351 GTGTTTGATG CGAAACCGAG ATGGGAGGTT GATAGGAAGC TTAATAAATT
 1401 GACAACTCGT GAGCAGGTGG AGAAAAATGT TCAGGAAACG AGAAGAAGGA
 1451 GTCAGAGTAG TCAGTTTAAA GCCCATGCGC AACGAGAATG GGAAATAAAA
 1501 ACAGGGTTAG ATTTTAATCA TTTTATAGGT GGTGATATCA ATAAGAAAGG
 1551 CACAGTAACA GGAGGGCATA GTCTAACCCG TGGTGATGTA CGGGTGATAC
 1601 AACAAACCTC GGCACCTGAT AAACATGGGG TTTATCAAGC GACAGTGGAA
 1651 ATTAATAAGC CTGATGGAAG TTGGGAGGTG AAAACGAAAA AAGGTGGGAA
 1701 AGTGATGACC AAGCACACCA TGTTCCCAA AGATTGGGAT GAGGCTAGAA
 1751 TTAGGGCTGA AGTTACTTCG GCTTGGGAAA GTAGAATAAT GCTTAAGGAT
 1801 AATAAATGGC AGGGTACAAG TAAATCGGGT ATTAATAATAG AAGGATTTAC
 1851 CGAACCTAAT AGAACAGCAT ATCCCATTTA TGAATAG

This corresponds to the amino acid sequence <SEQ ID 464; ORF46ng-1>:

1 LGISRKISLI LSILAVCLPM HAHASDLAND PFIRQVLDRO HFEPDGKGYHL
 51 FGSRGELAXR NGHIGLGNIO SHQLGHLMIQ QAAVEGNIGY IVRFSDHGKH
 101 FHSPFDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHPAD GYDGPQGGY
 151 PAPKGARDIY SYDIKVAQN IRLNLTNRS TGQRLADRFH NAGAMLTQGV
 201 GDGFKRATRY SPELDRSGNA AEAFTGTADI VKNIIGAAGE IVGAGDAVQG
 251 ISEGSNIAVM HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP
 301 NAAQIEAVS NIFMAAPIK GIGAVRGKYG LGGITAHVPK RSQMGATLAP
 351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYGKENI TSSTVPPSNG
 401 KNVKLADQRH PKTGVFPDGG GPFNFEXHVK YDTKLDIQL SGGGIPKAKP
 451 VEDAKPRWEV DRKLNKLTR EQVEKNVQET RRRSQSQFQ AHAEQREWENK
 501 TGLDFNHFIG GDINKKGTVT GGHSLTRGDV RVIQQTAPD KHGVYQATVE
 551 IKKPDGSWEV KTKKGGKVM THTMFPKDW EARIRAEVTS AWESRIMLKD
 601 NKWQTSKSG IKIEGFTEPN RTAYPIYE*

30 ORF46ng-1 and ORF46-1 show 94.7% identity in 227 aa overlap:

		10	20	30	40
orf46-1.pep		AVCLPMHAASXLANDSFIRQVLDROHFEPDGKGYHLFGSRGELAER			
orf46ng-1	LGISRKISLILSILAVCLPMHAASDLANDPFIRQVLDROHFEPDGKGYHLFGSRGELAXR				
		10	20	30	40
		50	60	70	80
orf46-1.pep	QSHIGLGKIQSHQLGNLMIQQAIAKGNIGYIVRFSDHGHEVHSPFDNHASHSDSDEAGSP				
orf46ng-1	NGHIGLGNIQSHQLGHLMIQQAIAVEGNIGYIVRFSDHGKHFKHSPFDNHASHSDSDEAGSP				
		70	80	90	100
		110	120	130	140
orf46-1.pep	VDGFSLYRIHWDGYEHPADGYDGPQGGGYPAPKGARDIYSYDIKVAQNIRLNLTNRS				
orf46ng-1	VDGFSLYRIHWDGYEHPADGYDGPQGGGYPAPKGARDIYSYDIKVAQNIRLNLTNRS				
		130	140	150	160
		170	180	190	200
orf46-1.pep	TGQRLADRFHNAGSMLTQGVGDGFKRATRYSPELDERSGNAEAFTGTADIVKNIIGAAGE				
orf46ng-1	TGQRLADRFHNAGAMLTQGVGDGFKRATRYSPELDERSGNAEAFTGTADIVKNIIGAAGE				
		190	200	210	220
		230	240	250	260
		270	280	290	300
orf46-1.pep	I				
orf46ng-1	IVGAGDAVQGISSEGSNIAVMHGLLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP				

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF46ng-1 shows 87.4% identity over a 486aa overlap with an ORF (ORF46a) from strain A of

N. meningitidis:

10 20 30 40 50 60

WO 99/24578

-276-

	orf46a.pep	LGISRKISLILSILAVCLPMHAHASDLANDSFIRQVLDROHFE PDGKYHLFGSRGELAER
	orf46ng-1	LGISRKISLILSILAVCLPMHAHASDLANDPFIRQVLDROHFE PDGKYHLFGSRGELAXR
5		10 20 30 40 50 60
	orf46a.pep	70 80 90 100 110 120
	orf46ng-1	SGHIGLGNIQSHQLGNLFIQQAIAKGNIGYIVRFS DHGHEVHSPFDNHASHSDSDEAGSP
10		70 80 90 100 110 120
	orf46a.pep	130 140 150 160 170 180
	orf46ng-1	VDGFSLYRIHWDGYEHHPADGYDGPQGGYPAPKGARDIYSYDIKGVAQNIRNLNTDNRS
15		130 140 150 160 170 180
	orf46a.pep	190 200 210 220 230 240
	orf46ng-1	TGQRLVDRFHNTGSMLTQGVGDGFKRATRYSPELD RSGNAAEFNGTADIVKNIIGAAGE
20		190 200 210 220 230 240
	orf46a.pep	250 260 270 280 290 300
	orf46ng-1	IVGAGDAVQGISSEGSNIAVMHGLGLLSTENKMARINDLADMAQLKDYAAAAIRDWAVQNP
25		250 260 270 280 290 300
	orf46a.pep	310 320 330 340 350 360
	orf46ng-1	NAAQGIEAVSNIFTAVIPVKIGAVRGKYGLGGITAH PVKRSQMG EIALPKGKSAVSDNF
30		310 320 330 340 350 360
	orf46a.pep	370 380 390 400 410 420
	orf46ng-1	ADAAYAKYPSPYHSRNI RSNLEQRYKENITSSTVPPSNGKNVKLAN KRHPKTKVPFDGK
35		370 380 390 400 410 420
	orf46a.pep	430 440 450 460 470
	orf46ng-1	GFPNFEKDVKYDTRINTAVPQVN----PIDEPVFN--PKGSVGS AHSWSITARIQYAKLP
40		430 440 450 460 470
	orf46a.pep	480 490 500 510 520 530
	orf46ng-1	RQGRIRYIPPKNYSAPLPKGPNNGYLDKFGNEWT KGPSRTKGQEFWDVQLSKTGREQ
45		480 490 500 510 520 530
	orf46a.pep	540 550 560 570 580 590
	orf46ng-1	QETRRRSQSSQFKAHAQREWENKTGLDFNHFIGGDINKKGTVTGGHSLTRGDV RVIQOTS
50		540 550 560 570 580 590

The complete length ORF46a DNA sequence <SEQ ID 465> is:

55	1	TTGGGCATT	CCCGCAAAT	ATCCCTTATT	CTGTCCATAC	TGGCAGTGTG
	51	CCTGCCGATG	CATGCACACG	CCTCAGATT	GGCAAACGAT	TCTTTTATCC
	101	GGCAGTTCT	CGACCGTCAG	CATTTCGAAC	CCGACGGGAA	ATACCACCTA
	151	TTCGGCAGCA	GGGGGGAAC	TGCCGAGCGC	AGCGGTCATA	TCGGATTGGG
	201	AAACATACAA	AGCCATCAGT	TGGGCAACCT	GTTTCATCCAG	CAGGCGGCCA
	251	TTAAAGGAAA	TATCGGCTAC	ATTGTCCGCT	TTTCCGATCA	CGGGCAGCAA
60	301	GTCCATTCCC	CCTTCGACAA	CCATGCCTCA	CATTCCGATT	CTGATGAAGC
	351	CGGTAGTCCC	GTTGACGGAT	TCAGCCTTTA	CCGCATCCAT	TGGGACGGAT
	401	ACGAACACCA	TCCCGCCGAC	GGCTATGACG	GGCCACAGGG	CGGCGGCTAT
	451	CCCGCTCCCA	AAGGCGCGAG	GGATATATAC	AGCTACGACA	TAAAGGCGT
	501	TGCCCAAAT	ATCCGCCTCA	ACCTGACCGA	CAACCGCAGC	ACCGGACAA
65	551	GGCTTGTCGA	CCGTTTCCAC	AATACCGGTA	GATGCTGAC	GCAAGGAGTA
	601	GGCGACGGAT	TCAAACGGCG	CACCCGATAC	AGCCCCGAGC	TGGACAGATC
	651	GGGCAATGCC	GCCGAAGCTT	TCAACGGCAC	TGCAGATATC	GTCAAAAACA
	701	TCATCGGCGC	GGCAGGAGAA	ATTGTCGGCG	CAGGCGATGC	CGTGCAGGGT
	751	ATAAGCGAAG	GCTCAAACAT	TGCTGTTATG	CACGGCTTGG	GTCTGCTTTC
70	801	CACCGAAAC	AAGATGGCGC	GCATCAACGA	TTTGGCAGAT	ATGGCGCAAC

851 TCAAAGACTA TGCCGCAGCA GCCATCCGCG ATTGGGCAGT CCAAAACCCC
 901 AATGCCGCAC AAGGCATAGA AGCCGTCAGC AATATCTTTA CGGCAGTCAT
 951 CCCCCTCAAA GGGATTGGAG CTGTTCCGGG AAAATACGGC TTGGGCGGCA
 1001 TCACGGCACA TCCTGTCAAG CGGTCGAGA TGGGCGAGAT CGCATTGCCG
 1051 AAAGGGAAAT CCGCCGTCAG CGACAATTTT GCCGATGCGG CATA CGCCAA
 1101 ATACCCGTCC CCTTACCATT CCGAAATAT CCGTTCAAAC TTGGAGCAGC
 1151 GTTACGGCAA AGAAAACATC ACCTCCTCAA CCGTGCCGCC GTCAAACGGA
 1201 AAGAATGTGA AACTGGCAA CAAACGCCAC CCGAAGACCA AAGTGCCGTT
 1251 TGACGGTAAA GGGTTTCCGA ATTTTGAAAA AGACGTAAAA TACGATACGA
 1301 GAATTAATAC CGCTGTACCA CAAGTGAATC CTATAGATGA ACCCGTCTTT
 1351 AATCCTAAAG GTTCTGTCCG ATCGGCTCAT TCTTGGTCTA TAACTGCCAG
 1401 AATTCAATAC GCAAAATTAC CAAGGCAAGG TAGAATCAGA TATATCCAC
 1451 CTAAAAATTA CTCTCCTTCA GCACCGCTAC CAAAGGACC TAATAATGGA
 1501 TATTTGGATA AATTTGGTAA TGAATGGACT AAAGGTCCAT CAAGAACTAA
 1551 AGGTCAAGAA TTTGAATGG ATGTTCAATT GTCTAAAACA GGAAGAGAGC
 1601 AACTTGGATG GGCTAGTAGG GATGGTAAGC ATTTAAATAT ATCAATTGAT
 1651 GGAAAGATTA CACACAAATG A

This corresponds to the amino acid sequence <SEQ ID 466>:

1 LGISRKISLI LSILAVCLPM HAHASDLAND SFIROVLDRQ HFEPDGKYHL
 51 FGSRGELAER SGHIGLGNLQ SHQLGNLFIQ QAAIKGNIGY IVRFSDHGHE
 101 VHSPFDNHAS HSDSDEAGSP VDGFSLYRIH WDGYEHHPAD GYDGPQGGGY
 151 PAPKGARDIY SYDIKGVAQN IRLNLTDNRS TGQRLVDRFH NTGSMLTQGV
 201 GDGFKRATRY SPELDRSGNA AEFNGTADI VKNIIGAAGE IVGAGDAVQG
 251 ISEGSNIAMV HGLGLLSTEN KMARINDLAD MAQLKDYAAA AIRDWAVQNP
 301 NAAQGLEAVS NIFTAVIPVK GIGAVRGKYG LGGITAHVPK RSQMGELALP
 351 KGKSAVSDNF ADAAYAKYPS PYHSRNIRSN LEQRYKENI TSSTVPPSNG
 401 KNVKLANKRH PKTKVPFDGK GFPNFEKDVK YDTRINTAVP QVNPIDEPVF
 451 NPKGSVGSAA SWSITARIQY AKLPRQGRIR YIPPKNYSPS APLPKGPNNNG
 501 YLDKFGNEWT KGPSRTKGQE FEWDVQLSKT GREQLGWASR DGKHLNISID
 551 GKITHK*

Based on this analysis, including the presence of a RGD sequence in the gonococcal protein, typical of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 56

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 467>:

1 ATGAATATTC ACACCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCTCT
 51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG
 101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
 151 TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCCTTTCGT
 201 CAAAATTGCC GGCCTATTGG CGTTTGGCT GCGGTTTGT TTTGACGGGC
 251 TGATGATGGT GATCCAATC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
 301 AACCTCGTCC CCTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC
 351 CGGGCTG...

This corresponds to the amino acid sequence <SEQ ID 468; ORF48>:

1 MNIHTLLSKQ WTLPPFLPKR LLLSLILLA PNAVFWLAL LTATARPIVN
 51 LDYLPALLI ALPWRVFKIA GVLAFLAVL FDGLMMVIQL FPFMDLIGAI
 101 NLVPFILTAP APYQIMTGL...

Further work revealed the complete nucleotide sequence <SEQ ID 469>:

1 ATGAATATTC ACACCCTGCT CTCCAAACAA TGGACGCTGC CGCCATTCTCT
 51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTTGCC CCCAATGCGG
 101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCCGCCC GATTGTCAAT
 151 TTGGACTATC TTCCCGCCGC GCTGCTGATC GCCCTGCCTT GGCCTTTCGT
 201 CAAAATTGCC GGCCTATTGG CGTTTGGCT GCGGTTTGT TTTGACGGGC
 251 TGATGATGGT GATCCAATC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
 301 AACCTCGTCC CCTTCATCCT GACCGCCCCC GCCCCTTATC AGATAATGAC
 351 CGGGCTGTTG CTGCTGTATA TGCTGGCGAT GCCGTTTGTG TTGCAGAAAG

5
10
15
20

```

401 CCGCCGCCAA AACCGACTTC CGGCACATTG CCGTCTGCGC CGCCGTTGTG
451 GCGGCAGCCG GCTATTTTAC CGGCCATTG AGTTACTACG ACCGGGGTGC
501 GATGGCCAAT ATCTTCGGCG CAAACAATT CTACTACGCC AAAAGTCAGG
551 CGATGCTCTA CACCGTCAGC CAGAATGCCG ACTTTATTAC CGCCGGCCTG
601 GTCGATCCCG TCTTCTCCC CTTGGGCAAT CAACAGCGTG CGGCCACGCA
651 TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT
701 GGGGGCTGCC GGCCAATCCC GAACTTCAAA ACGCCACTTT TGCCAAACTG
751 CTGGCGCAAA AAGACCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCCTT
801 CATCGGCGCG ACGGTCGAAG GCGAAATGCG CGAACTGTGT GCCTACGGCG
851 GTTTGCGCGG GTTCGCACTG CGCCGCGCGC CGACGAAAA ATTTGCCCGC
901 TGCCTCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCGATGCA
951 CGGCGCGGCG AGTTCGCTTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG
1001 GCTTTCAAGA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAAACCTGC
1051 GCCATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTGCGCG AAGTGTCCGG
1101 ATTTTCAAA AAACACGACA AGGGACTGTT TTAAGGATG ACCTGACCA
1151 GCCACGCCGA CTATCCCGAA TCCGACATT TCAACCACAG GCTCAAATGC
1201 ACCGAATATG GCCTGCCCGC CGAAACCGAC CTCTGCCGCA ATTTACAGCT
1251 GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCCAA CGCCCGGAAA
1301 TGAAGGCAC GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCCCGTCGGC
1351 AACCTCAATG AAACCTCCG CTACCTCAA CAGGGGCACG TCGCTGGCT
1401 GAACCTCAA ATCAATAA

```

This corresponds to the amino acid sequence <SEQ ID 470; ORF48-1>:

25
30

```

1 MNIHTLLSKQ WTLPPFLPKR LLLSLILLA PNAVFWLAL LTATARPIVN
51 LDYLPALLI ALPWRVFKIA GVLAFLAVL FDGLMMVIQL FPFMDLIGAI
101 NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAKTD F RHIAVCAAVV
151 AAAGYFTGHL SYDRGRMAN IFGANFYA KSQAMLYTVS QNADFITAGL
201 VDPVFLPLGN QORAATHLNE PKSQKILFIV AESWGLPANP ELQATFAKL
251 LAQKDRFSVW ESGSFPIGA TVEGEMREL CAYGGLRGFAL RRAPDEKFA
301 CLPNRLKQEG YATFAMHGAG SSLYDRFSW PRAGFQEIKT AENLIGKTC
351 AIFGGVCDSE LFGEVSAFFK KHDKGLFYWM TLTSHADYPE SDIFNHLKLC
401 TEYGLPAETD LCRNFSLHTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPVPG
451 NLNETFRYLK QGHVAVLNFK IK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

35 ORF48 shows 94.1% identity over a 119aa overlap with an ORF (ORF48a) from strain A of *N. meningitidis*:

40
45
50

```

          10      20      30      40      50      60
orf48.pep MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWLALLLTATARPIVNDYLPALLI
          10      20      30      40      50      60
40 orf48a  MNIHTLLSKQWTLPPFLPKRLLLSLLILLXPNVFWLALLLTATARPIVNLXYLPALLI
          10      20      30      40      50      60

          70      80      90      100     110     119
orf48.pep ALPWRVFKIAGVLAFLAVLFDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGL
          70      80      90      100     110     120
45 orf48a  ALPWRXVKIXGLAXWLAVLFDGLMMVIQLFPFMDLIGAINLVPFXTAPALYQIMTGLL
          70      80      90      100     110     120

          130     140     150     160     170     180
50 orf48a  LLYMLAMPFVLQKAAKTD F RHIAACA VVVAAGYFTGHL SXDRGRMAN IFGANFYA

```

The complete length ORF48a nucleotide sequence <SEQ ID 471> is:

55
60

```

1 ATGAATATTC ACACCCTGCT CTCCAACAA TGGACGCTGC CGCCATTCCT
51 GCCGAAACGG CTGCTGCTGT CCCTGCTGAT ACTGCTNNCC CCCAATGCGG
101 TGTTTTGGGT TTTGGCACTG CTGACCGCCA CCGCCGCCC GATTGTCAAT
151 TTGGANTACC TTCCCGCCG CTTGCTGATC GCCCTGCCTT GCGGTNTCGT
201 CAAAATTGNC GCGGTATTGG CGTNTTGGCT GCGGTTTGT TTTGACGGGC
251 TGATGATGGT GATCCAATC TTCCCTTTTA TGGATCTCAT CGGCGCCATC
301 AACCTCGTCC CCTTCATCNT GACCGCCCC GCCCTTATC AGATAATGAC
351 CGGGCTGTTA CTGCTGTATA TGCTGGCGAT GCCGTTTGTG TTGCAGAAAG
401 CCGCCGCCAA AACCGACTTC CGACACATTG CCGCCTGTGC CGCCGTTGTG
451 GTGGCAGCCG GCTATTTTAC CGGCCATTG AGTTANTACG ACCGGGGGCG

```

-279-

501 GATGGCCAAT ATCTTCGGCG CAAACAACCTT CTATTACGCC AAAAGTCAGG
 551 CGATGCTCTA CACCGTCAGC CAGAAATGCC ACTTTATTAC CGCCGCGCTG
 601 GTCGATCCCG TCTTCTCTCC CTTGGGCAAT CAACAGCGTG CCGCCACGCA
 651 TCTGAACGAG CCGAAATCTC AAAAAATCCT CTTTATCGTC GCCGAATCTT
 701 GGGGGCTGCC GGCCAATCCC GAACTTCAAA ACGCCACTTT TGCCAACTG
 751 CTGGCGCAA AAGANCGTTT TTCGGTTTGG GAAAGCGGCA GTTTTCCTCT
 801 CATCGGCGCG ACGATCGAAG GCGAAATGCG CGAACTGTGT GCCTACGGCG
 851 GTTTCGCGCG GTTCGCACTG CGCCGCGCGC CCGACGAAAA ATTTGCCCGC
 901 TGCCTCCCCA ACCGTTTGAA ACAAGAAGGT TACGCCACCT TTGCGATGCA
 951 CGGCGCGGGG AGTTTCGCTTT ACGACCGCTT CAGCTGGTAT CCGAGGGCGG
 1001 GCTTTCAGA AATCAAAACC GCCGAAAACC TGATCGGTAA AAAAACCTGC
 1051 GCCATTTTCG GCGGCGTGTG CGACAGCGAG CTGTTTCGGC AAGTGTCCGC
 1101 ANTTTTCAAA AAACACGACA AGGGACTGTT TTAAGGATG ACGCTGACCA
 1151 GCCACGCCGA CTATCCCGAA TCNGACATTT TCAACCACAG GCTCAAATGC
 1201 ACCGAATATG GCCTGCCCGC CGAAACCGAC NTCTGCCGCA ATTTAGCCT
 1251 GCACACCCAA TTCTTCGACC AACTGGCGGA TTTGATCCAA CGCCCGGAAA
 1301 TGAAAGGCAC GGAAGTCATC ATCGTCGGCG ACCATCCGCC GCCCGTCGGC
 1351 AACCTCAATG AAACCTTCCG CTACCTCAA CAGGGGCACG TCGNCTGGCT
 1401 GAACTTCAA ATCAATAA

20 This encodes a protein having amino acid sequence <SEQ ID 472>:

1 MNIHTLLSKQ WTLPPFLPKR LLLSLILLX PNAVFVWLAL LTATARPIVN
 51 LXLYPAALLI ALPWRXVKIX GVLAXWLAVL FDGLMMVIQL FPFMDLIGAI
 101 NLVPFXTAP ALYQIMTGLL LLYMLAMPFV LQKAAKTDF RHIAACAADV
 151 VAAGYFTGHL SXDRGRMAN IFGANNFYA KSQAMLYTVS QNADFITAGL
 201 VDPVFLPLGN QQRAATHLNE PKSQILFIV AESWGLPANP ELQNAFPAKL
 251 LAQKXRFVSW ESGSFPIGA TIEGEMREL CAYGGLRGFAL RRAPEKFAF
 301 CLPNRLKQEG YATFAMHGAG SLYDRFSWY PRAGFQEIKT AENLIGKKT
 351 AIFGGVCDSE LFGEVSAXFK KHDKGLFYWM TLTSHADYE SDIFNHLKLC
 401 TEYGLPAETD XCRNFSLTQ FFDQLADLIQ RPEMKGTEVI IVGDHPPPVG
 451 NLNETFRYLK QGHVXWLNFK IK*

ORF48a and ORF48-1 show 96.8% identity in 472 aa overlap:

		10	20	30	40	50	60
orf48a.pep		MNIHTLLSKQWTLPPFLPKRLLLSLILLX	PNAVFVWLALLTATARPIVNLXLYPAALLI				
35	orf48-1	MNIHTLLSKQWTLPPFLPKRLLLSLILLAPNAVFVWLALLTATARPIVNLQDYLPAALLI					
		10	20	30	40	50	60
	orf48a.pep	ALPWRXVKIXGVLAXWLAVLFDGLMMVIQLFPFMDLIGAINLVVPFXTAPALYQIMTGLL					
40	orf48-1	ALPWRXVKIXGVLAXWLAVLFDGLMMVIQLFPFMDLIGAINLVVPFXTAPALYQIMTGLL					
		70	80	90	100	110	120
	orf48a.pep	LLYMLAMPFVLQKAAKTDFRHIAACAADVVAAGYFTGHL SXDRGRMANIFGANNFYA					
45	orf48-1	LLYMLAMPFVLQKAAKTDFRHIAACAADVVAAGYFTGHL SXDRGRMANIFGANNFYA					
		130	140	150	160	170	180
	orf48a.pep	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATHLNEPKSQILFIVAESWGLPANP					
50	orf48-1	KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATHLNEPKSQILFIVAESWGLPANP					
		190	200	210	220	230	240
	orf48a.pep	ELQNAFPAKLLAQKXRFVSWESGSFPIGATIEGEMRELCAYGGLRGFALRRAPDEKFAF					
55	orf48-1	ELQNAFPAKLLAQKXRFVSWESGSFPIGATIEGEMRELCAYGGLRGFALRRAPDEKFAF					
		250	260	270	280	290	300
	orf48a.pep	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKTAEENLIGKKTCAIFGGVCDSE					
60	orf48-1	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKTAEENLIGKKTCAIFGGVCDSE					
		310	320	330	340	350	360
	orf48a.pep	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKTAEENLIGKKTCAIFGGVCDSE					
65	orf48-1	CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKTAEENLIGKKTCAIFGGVCDSE					

-280-

		370	380	390	400	410	420
	orf48a.pep	LFG	EV	SAX	FKK	HKD	KGLFY
5	orf48-1	LFG	EV	SAX	FKK	HKD	KGLFY
		370	380	390	400	410	420
	orf48a.pep	FFD	QLA	DLI	QRP	EMK	GTE
10	orf48-1	FFD	QLA	DLI	QRP	EMK	GTE
		430	440	450	460	470	

Homology with a predicted ORF from *N.gonorrhoeae*ORF48 shows 97.5% identity over a 119aa overlap with a predicted ORF (ORF48ng) from *N.*15 *gonorrhoeae*:

	orf48.pep	MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFVWLALLTATARPIVNL DYLP AALLI	60
		:	
	orf48ng	MNIHALLSEQWTLPPFLPKRLLLSLLILLAPNAVFVWLALLTATARPIVNL DYLP AALLI	60
20	orf48.pep	ALPWR FVKIAGVLA FVLAFLVDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGL	119
	orf48ng	ALPWR FVKIAGVLA FVLAFLVDGLMMVIQLFPFMDLIGAINLVPFILTAPAPYQIMTGL	120

The ORF48ng nucleotide sequence <SEQ ID 473> was predicted to encode a protein having amino acid sequence <SEQ ID 474>:

25	1	MNIHALLSEQ	WTLP	PF	PKR	LLLS	LLILLA	PNAVFVWLAL	LTATARPIVN
	51	LDYLP	AALLI	ALPWR	FVKIA	GVLA	FVLAFL	VDGLMMVIQL	FPFMDLIGAI
	101	NLVPF	ILTAP	APYQ	IMTGLL	LLYML	AMPFV	LQKA	AVKTDF
	151	AAARY	FTGPF	ELLRT	GGRWQ	YVQHR	RLLS	GSRAS	FRRRQ
	201	PYAS	MNGG						

30 Further work identified the complete gonococcal DNA sequence <SEQ ID 475>:

	1	ATGAATATTC	ACGCCCTGCT	CTCCGAACAA	TGGACGCTGC	CGCCATTCCCT
	51	GCCGAAACGG	CTGCTGCTGT	CCCTGCTGAT	ACTGCTGGCC	CCCAATGCCG
	101	TGTTTTGGGT	TTTGGCACTG	CTGACCGCCA	CCGCCCCGCC	GATTGTCAAT
	151	TTGGACTACC	TTCCCGCCGC	GCTGCTGATC	GCCCTGCCTT	GGCGTTTCGT
35	201	CAAATTTGCC	GGCGTATTGG	CGTTTTGGCC	GGCGGTTTTG	TTTGACGGGC
	251	TGATGATGGT	GATCCAATC	TTCCCTTTTA	TGGACCTCAT	CGGCGCCATC
	301	AACCTCGTCC	CCTTCATCCT	GACGCGCCCC	GCCCTTATC	AGATAATGAC
	351	CGGGCTGTG	CTGCTGTATA	TGCTGGCGAT	GCCGTTTGTG	TTGCAAAAAG
	401	CCGCCGTCAA	AACCGACTTC	CGACACATTG	CCGTCTGTGC	CGCCGTTGTG
40	451	GCGGCAGCCG	GCTATTTTCA	CGGCCATTG	AGTTACTACG	ACCGGGGGCG
	501	GATGGCCAAT	ATCTTCGGCG	CAAACAACCT	CTATTACGCC	aAAAGTCAGG
	551	CGATGCTCTA	CACCGTCAGC	CAGAATGCCG	ACTTTATTAC	CGCCGgcctg
	601	GTCGACCCCG	TCTTCCTCCC	CTTGGGCAAT	CAGCAGCGTG	CGGCCACGCG
	651	GCTGAGTGAG	CCGAAATCTC	AAAAATCCT	CTTTATCGTC	GCCGAATCTT
45	701	GGGGGCTGCC	GGGCAATCCC	GAGCTTCAAA	ACGCCACTTT	TGCCAAACTG
	751	CTGGCGCAAA	AAGACCGTTT	TTCGGTTTGG	GAAAGCGGCA	GTTTTCCTT
	801	CATCGGCGCG	ACGGTCAAG	GCGAAATGCG	CGAATTGTGC	GCCTACGGCG
	851	GTTTGCGCGG	GTTTCGCACTG	CGCGCGCGC	CCGACGAAAA	ATTTGCCCGC
	901	TGCCTCCCA	ACCGTTTGAA	ACAAGAAGGT	TACGCCACCT	TTGCGATGCA
50	951	CGGCGCGGGT	AGTTCGCTTT	ACGACCGCTT	CAGCTGGTAT	CCGAGGGCGG
	1001	GCTTCAAAA	AATCAAAACC	GCCGAAAACC	TGATCGGTAA	AAAAACCTGC
	1051	GCCATTTTCG	GCGGCGGTGTG	CGACAGCGAG	CTGTTCCGGC	AAGTGTCCGGC
	1101	ATTTTTCAAA	AAACACGACA	AGGGACTGTT	TTACTGGATG	ACGCTGACCA
	1151	GCCACGCCGA	CTATCCCGAA	TCCGACATTT	TCAACCACAG	GCTCAAATGC
55	1201	ACCGAATACG	GCCTGCCCGC	CGAAACCGAC	CTCTGCCGCA	ATTTACGCTT
	1251	GCACACCCAA	TtcttcgACC	AACTGGCGGA	TTTGATCCGA	CGCCCCGAAA
	1301	TGAAAGGCAC	GGAAGTCATC	ATCGTCGGCG	ACCATCCGCC	CGCCGTCGGC
	1351	AACCTCAATG	AAACCTTCCG	CTACCTCAAA	CAGGGACACG	TCGCCTGGCT
	1401	GCAC	TTCAAA	ATCA	ATAA	

60 This encodes a protein having amino acid sequence <SEQ ID 476; ORF48ng-1>:

```

      1  MNIHALLSEQ WTLPPFLPKR LLLSLILLA PNAVFWLAL LTATARPVIV
    51  LDYLPALLI ALPWRVFKIA GVLAFWPAVL FDGLMMVIQL FPFMDLIGAI
   101  NLVPFILTAP APYQIMTGLL LLYMLAMPFV LQKAAVKTD RHIIVCAAVV
   151  AAAGYFTGHL SYDRGRMAN IFGANNFYA KSQAMLYTVS QNADFITAGL
   201  VDPVFLPLGN QQRAATRLSE PKSOKILFIV AESWGLPGNP ELQATFAKL
   251  LAQKDRFSVW ESGSFPIGA TVEGEMREL CAYGGLRGFAL RRAPDEKFA
   301  CLPNRLKQEG YATFAMHGAG SSLYDRFSWY PRAGFQIKT AENLIGKTC
   351  AIFGGVCDSE LFGEVSAFFK KHDKGLFYWM TLTSHADYPE SDIFNHLKLC
   401  TEYGLPAETD LCRNFSLHTQ FFDQLADLIR RPEMKGTEVI IVGDHPPVVG
   451  NLNETFRYLK QGHVAWLHFK IK*

```

ORG48ng-1 and ORF48-1 show 97.9% identity in 472 aa overlap:

```

      10      20      30      40      50      60
orf48-1.pep  MNIHTLLSKQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPVIVNLDYLPALLI
15 orf48ng-1  MNIHALLSEQWTLPPFLPKRLLLSLLILLAPNAVFWVLALLTATARPVIVNLDYLPALLI
      10      20      30      40      50      60
      70      80      90     100     110     120
orf48-1.pep  ALPWRVFKIAGVLAFWLAVLFDGLMMVIQLFPFMDLIGAINLVVPFILTAPAPYQIMTGLL
20 orf48ng-1  ALPWRVFKIAGVLAFWPAVLFDGLMMVIQLFPFMDLIGAINLVVPFILTAPAPYQIMTGLL
      70      80      90     100     110     120
      130     140     150     160     170     180
orf48-1.pep  LLYMLAMPFVLQKAAAKTDFRHIIVCAAVVAAAGYFTGHLSSYDRGRMANIFGANNFYA
25 orf48ng-1  LLYMLAMPFVLQKAAVKTD RHIIVCAAVVAAAGYFTGHLSSYDRGRMANIFGANNFYA
      130     140     150     160     170     180
      190     200     210     220     230     240
orf48-1.pep  KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATHLNEPKSOKILFIVAESWGLPANP
30 orf48ng-1  KSQAMLYTVSQNADFITAGLVDPVFLPLGNQQRAATRLSEPKSOKILFIVAESWGLPGNP
      190     200     210     220     230     240
      250     260     270     280     290     300
orf48-1.pep  ELQATFAKLLAQKDRFSVWESGSFPIGATVEGEMRELCAYGGLRGFALRRAPDEKFA
35 orf48ng-1  ELQATFAKLLAQKDRFSVWESGSFPIGATVEGEMRELCAYGGLRGFALRRAPDEKFA
      250     260     270     280     290     300
      310     320     330     340     350     360
orf48-1.pep  CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQEIKAENLIGKKTCAIFGGVCDSE
45 orf48ng-1  CLPNRLKQEGYATFAMHGAGSSLYDRFSWYPRAGFQIKIKAENLIGKKTCAIFGGVCDSE
      310     320     330     340     350     360
      370     380     390     400     410     420
orf48-1.pep  LFGEVSAFFKKHDKGLFYWMTLTSHADYPESDI FNHRLKCTEYGLPAETDLCRNFSLHTQ
50 orf48ng-1  LFGEVSAFFKKHDKGLFYWMTLTSHADYPESDI FNHRLKCTEYGLPAETDLCRNFSLHTQ
      370     380     390     400     410     420
      430     440     450     460     470
orf48-1.pep  FFDQLADLIQRPEMKGTEVIIVGDHPPVGNLNETFRYLKQGHVAWLNFKIKX
55 orf48ng-1  FFDQLADLIIRPEMKGTEVIIVGDHPPVGNLNETFRYLKQGHVAWLHFKIKX
      430     440     450     460     470

```

Based on this analysis, including the presence of a putative leader sequence (double-underlined)

60 and two putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 57

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 477>:

```

      1  ..GTGAGCGGAC GTTACCGCGC TTTGGATCGC GTTTCACAAA TCATCATCGT
      51  TACTTTGAGT ATCGCCACGC TTGCCGCCGC CGGCATCGCT ATGTCGCGCG
5      101  GTATGCAGAT GCAGTCCGAT TTTATCGAGC CGACACCGTG GACGCTTGCC
      151  GGTTTGGGCT TCCTGATCGC GCTGATGGGC TGGATGCCCG CGCCGATTGA
      201  AATTTCCGCC ATCAATTCTT TGTGGGTAAC CGAAAAACAA CGCATCAATC
      251  CTTCCGAATA CCGCGACGGG ATTTTGAAT TCAACGTCGG TTATATCGCC
      301  AGTGCGGTTT TGGCTTTGGT TTCTTGCA CTGGGCGC.G TAGCGCCGAA
10     351  CGGCAACGGC GA.ACAGTGC AGATGGCGGG CGGCAATAT AACGGGCAAT
      401  TGATCAATAT GTACGCC..

```

This corresponds to the amino acid sequence <SEQ ID 478; ORF53>:

```

      1  ..VSGRYRALDR VSKIIIVTLS IATLAAAGIA MSRGMQMOSD FIEPTPWTLA
      51  GLGFLIALMG WMPAPIEISA INSLWVTEKQ RINPSEYRDG IFEFNVGYIA
15     101  SAVLALVFLA LGXVAPNGNG XTVMAGGKY NQQLINMYA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 479>:

```

      1  ATGTCCGAAC AACATATTTT GACTTGGAAG AGTAAATCA ACGCATTGGG
      51  TCCGGGGATC ATGATGGCTT CGGCGGCGGT CGGCGGTTTC CACCTGATTG
      101  CCTCGACGCA GCGGGCGCGC CTTTACGGCT GGCAGATCGC GCTCATCATC
20     151  ATCCTGACCA ACCTCTTCAA ATACCGTTT TTCCGCTTCA GCGCGCATT
      201  CACGCTGGAC ACGGGCAAGA GCCTGATTGA AGGTTATGCC GAGAAAAGCC
      251  GCGTTTATTT GTGGGTATTC CTGATTTGT GCATCCTCTC CGCCACGATT
      301  AACGCGGGCG CGGTCCGCAT TGTAACCGCC GCCATCGTCA AAATGGCGAT
      351  TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCTTG ATTATGCGAT
25     401  CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT GGATCGCGTT
      451  TCCAAATCA TCATCGTTAC TTTGAGTATC GCCACGTTG CCGCCGCCGG
      501  CATCGCTATG TCGCGCGGTA TGCAGATGCA GTCCGATTTT ATCGAGCCGA
      551  CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT GATGGGCTGG
      601  ATGCCCGCGC CGATTGAAAT TTCCGCCATC AATTCTTTGT GGGTAACCGA
30     651  AAAACAACGC ATCAATCCTT CCGAATACCG CGACGGGATT TTTGATTTC
      701  ACGTCGGTTA TATCGCCAGT CGCGTTTGG CTTTGGTTTT CTTGCACTG
      751  GGCGCGTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA TGGCGGGCGG
      801  CAAATATATC GGGCAATTGA TCAATATGTA CGCCGTACC ATCGGCGGCT
      851  GGTGCGCCCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT GTACGGCAGC
35     901  ACGATTACCG TCGTGGACGG CTATGCCCGT GCCATTGCCG AACCCTGGCG
      951  CTGCTGCGC GGAAAAGACA AAACGGGCAA CGCCGAATTC TTGCGCTGGA
1001  ATATTTGGGT GGCGGGCAGC GGTGCGCGG TGATTTTCTG GTTGACGGC
1051  GTAATGGCGA ATCTGCTCAA ATTTGCGATG ATTGCCGCTT TTGTGTCCGC
1101  CCCTGTGTTT GCCTGGCTGA ATTACGTTT GGTAAAGGT GATGAAAAAC
40     1151  ACAAACCTAC ATCAGGTATG AATGCCCTTG CATTGGCAGG CTTGATTTAT
1201  CTGACCGGTT TTACCGTTTT GTTCTTATTG AATTGGCGG GAATGTTCAA
1251  ATGA

```

This corresponds to the amino acid sequence <SEQ ID 480; ORF53-1>:

```

      1  MSEQHISTWK SKINALGPGI MMASAAVGS HLIASTQAGA LYGWQIALII
45     51  ILTNLFKYPF FRFSAHYTLD TGKSLIEGYA EKSRYLWVF LILCILSATI
      101  NAGAVAIVTA AIVKMAIPSL MFDAGTVAAL IMASCLIIIV SGRYRALDRV
      151  SKIIIVTISI ATLAAAGIAM SRGMQMOSDF IEPTPWTLAG LGFLIALMGW
      201  MPAPIEISAI NSLWVTEKQ RINPSEYRDG IFEFNVGYIA AVLALVFLAL
50     251  GAFVQYNGE AVQMAGGKYI GQLINMYAVT IGGWSRPLVA FIAFACMYGT
      301  TITVVDGYAR AIAEPVRLLR GKDKTGNAEF FAWNIVVAGS GLAVIFWFDG
      351  VMANLLKFAM IAAFVSAPVF AWLNRYLVKG DEKHKLTSKM NALALAGLIY
      401  LTGFTVLFL NLAGMFK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

55 ORF53 shows 93.5% identity over a 139aa overlap with an ORF (ORF53a) from strain A of *N.meningitidis*:

-283-

```

                    10      20      30
orf53.pep          VSGRYALDRVSKIIIVTSLIATLAAAGIA
                    |||||
orf53a             AAIVKMAIPSLMFDAGTVAALIMASCLIIIVS
110      120      130      140      150      160
                    40      50      60      70      80      90
orf53.pep          MSRGMQMOSDFIEPTPWTLAGLGLIALMGWMPAPIEISAINSLWVTEKQRI
10      170      180      190      200      210      220
orf53a             MSRGMQMOSDFIEPTPWTLAGLGLIALMGWMPAPIEISAINSLWVTEKQRI
                    |||||
                    100      110      120      130      139
orf53.pep          IFEFNVGYIASAVLALVFLALGXVAPNGNGXTVQAGGKYNGQLINMYA
15      ||:|||||
orf53a             IFDFNVGYIASAVLALVFLALGAFVQYGNGEAVQAGGKYIGQLINMYAVTIGG
230      240      250      260      270      280
                    AFIAFACMYGTTITVVDGYARAIAEPVRLLRGKDKTGNAEFFAWN
20      290      300      310      320      330      340
orf53a             IAWNIWVAGSGLAVIFWFD

```

The complete length ORF53a nucleotide sequence <SEQ ID 481> is:

```

1  ATGTCCGAAC AACATATTTTC GACTTGGAAG AGTAAATCA ACGCATTGGG
51  ACCGGGGATT ATGATGGCTT CGGCGGCGGT CGGCGGTTCG CACCTGATTG
101 CCTCGACGCA GCGGGGCGCG CTTTACGGCT GGCAGATCGC GCTCATCATC
25 151 ATCCTGACCA ACCTCTTCAA ATACCCGTTT TTCCGCTTCA GCGCGCATT
201 CACGCTGGAC ACGGGCAAGA GCCTGATTGA AGGTTATGCC GAGAAAAGCC
251 GCGTTATTTT GTGGGTATTC CTGATTTTGT GCATCCTCTC CGCCACGATT
301 AACGCGGGCG CGGTCGCCAT TGTAAACGCC GCCATCGTCA AAATGGCGAT
351 TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG ATTATGGCAT
30 401 CCTGCCTGAT TATTTTGGTG AGCGGACGTT ACCGCGCTTT GGATCGCGTT
451 TCCAAAATCA TCATCGTTAC TTTGAGTATC GCCACGCTTG CCGCCGCCGG
501 CATCGCTATG TCGCGCGGTA TGCAGATGCA GTCCGATTTT ATCGAGCCGA
551 CACCGTGGAC GCTTGCCGGT TTGGGCTTCC TGATCGCGCT GATGGGCTGG
601 ATGCCCGCGC CGATTGAAAT TTCCGCCATC AATTCTTTGT GGGTAACCGA
35 651 AAAACAACGC ATCAATCCTT CCGAATACCG CGACGGGATT TTTGATTCA
701 ACGTCGGTTA TATCGCCAGT GCGGTTTGG CTTTGGTTTT CTTGATCATG
751 GGCGCGTTTG TGCAATACGG CAACGGCGAA GCAGTGCAGA TGGCGGGCGG
801 CAAATATATC GGGCAATTGA TCAATATGTA CGCCGTTACC ATCGGCGGCT
851 GGTGCGGCCC GCTGGTGGCG TTTATCGCGT TTGCCTGTAT GTACGGCAGC
40 901 ACGATTACCG TTGTGGACGG CTATGCCCGT GCCATTGCCG AACCCGTGCG
951 CCTGCTGCGC GGAAAAGACA AAACGGGCAA CGCCGAATTC TTTGCTGGA
1001 ATATTTGGGT GCGCGGCAGC GGTTTGGCGG TGATTTCTG GTTTGACGGC
1051 GTAATGGCGA ATCTGCTCAA ATTTGCGATG ATTGCCGCTT TTGTGTCCGC
1101 CCCTGTGTTT GCCTGGCTGA ATTACGTTT GGTCAAAGGT GATGAAAAAC
45 1151 ACAAACACAC ATCAGGTATG AATGCCCTTG CATTGGCAGG CTTGATTAT
1201 CTGACCGGTT TTACCGTTTT GTTCTTATTG AATTGGCGG GAATGTTCAA
1251 ATGA

```

This encodes a protein having amino acid sequence <SEQ ID 482>:

```

1  MSEQHISTWK SKINALGPGI MMASAAVGGG HLIASTQAGA LYGWQIALII
50 51  ILTNLFKYPF FRFSAHYTLD TGKSLIEGYA EKSRYLWVF LILCILSATI
101 NAGAVAIVTA AIVKMAIPSL MFDAGTVAAL IMASCLIIIV SGRYALDRV
151 SKIIIVTSLI ATLAAAGIAM SRGMQMOSDF IEPTPWTLAG LGFLIALMGW
201 MPAPIEISAI NSLWTEKOR INPSEYRDI FDFNVGYIAS AVLALVFLAL
251 GAFVQYGNGE AVQMAGGKYI GOLINMYAVT IGGWSRPLVA FIAFACMYGT
55 301 TITVVDGYAR AIAEPVRLR GKDKTGNAEF FAWNIVWAGS GLAVIFWFDG
351 VMANLLKFAM IAEVSAPEV AWLNRYLVKG DEKHLTSGM NALALAGLIY
401 LTGFTVLFLL NLAGMEFK*

```

ORF 53a shows 100.0% identity in 417 aa overlap with ORF53-1:

```

                    10      20      30      40      50      60
60 orf53a.pep      MSEQHISTWKSINALGPGIMMASAAVGGSHLIASTQAGALYGWQIALIIILTNLFKYPF
                    |||||
orf53-1            MSEQHISTWKSINALGPGIMMASAAVGGSHLIASTQAGALYGWQIALIIILTNLFKYPF
                    10      20      30      40      50      60
65      70      80      90      100      110      120

```

-284-

5	orf53a.pep	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAIVKMAIPSL
	orf53-1	FRFSAHYTLDTGKSLIEGYAEKSRVYLWVFLILCILSATINAGAVAIVTAIVKMAIPSL
10	orf53a.pep	MFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTSLIATLAAAGIAMSRGMQMSDF
	orf53-1	MFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTSLIATLAAAGIAMSRGMQMSDF
15	orf53a.pep	IEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGIFDFNVGYIAS
	orf53-1	IEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDGIFDFNVGYIAS
20	orf53a.pep	AVLALVFLALGAFVQYGNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVAFIAFACMYGT
	orf53-1	AVLALVFLALGAFVQYGNGEAVQMAGGKYIGQLINMYAVTIGGWSRPLVAFIAFACMYGT
25	orf53a.pep	TITVVDGYARAIAPVRLRLRGKDKTGNAEFFAWNIVAGSGLAVIFWFDGVMANLLKFAM
	orf53-1	TITVVDGYARAIAPVRLRLRGKDKTGNAEFFAWNIVAGSGLAVIFWFDGVMANLLKFAM
30	orf53a.pep	IAAFVSAPVFAWLNRYLVKGDEKHKLTSGMNALALAGLIYLTGFTVLFLLNLAMGFKX
	orf53-1	IAAFVSAPVFAWLNRYLVKGDEKHKLTSGMNALALAGLIYLTGFTVLFLLNLAMGFKX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF53 shows 92.1% identity over a 139aa overlap with a predicted ORF (ORF53ng) from *N. gonorrhoeae*:

40	orf53.pep	VSGRYRALDRVSKIIIVTSLIATLAAAGIA	30
	orf53ng	AAIVKMAIPSLMFDAGTVAALIMASCLIIIVSGRYRALDRVSKIIIVTSLIATLAAAGIA	91
45	orf53.pep	MSRGMQMSDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	90
	orf53ng	MSRGMQMPDFIEPTPWTLAGLGFLIALMGWMPAPIEISAINSLWVTEKQRINPSEYRDG	151
50	orf53.pep	IFEFNVGYIASAVLALVFLALGXVAPNGNGXTVQMAGGKYNGQLINMYA	139
	orf53ng	IFDFNVGYIASAVLALVFLALGAFVQYGNGEAVQMGGGKYIGQLINMYAVTIGGGSRLPV	211

An ORF53ng nucleotide sequence <SEQ ID 483> was predicted to encode a protein having amino acid sequence <SEQ ID 484>:

55	1	MPKKSCVYLW VFLILCIASA TINAGAVAIV TAAIVKMAIP SLMFDAGTVA
	51	ALIMASCLII LVSGRYRALD RVSKIIIVTL SIATLAAAGI AMSRGMQMP
	101	DFIEPTWTL AGLGFLIALM GWMPAPIEIS AINSLWVTEK QRINPSEYRD
	151	GIFDFNVGYI ASAVLALVFL ALGAFVQYGN GEAVQMGGGK YIGQLINMYA
	201	VTIGGGSRLP VAFIAFACMY GAASTVVDGY ARAIAEPVRL LRGKDKTARP
	251	IVLLEKLGR HRFGRDFLV*

Further analysis revealed further partial DNA gonococcal sequence <SEQ ID 485>:

60	1	..aagaAAAGCT GCGTTTATTT GTGGGTTTTT TTGATTTTGT GTATCGCCTC
	51	CGCCACGATT AACGCGGGCG CGGTCGCCAT TGTAACCGCC GCCATCGTCA
	101	AAATGGCGAT TCCCTCGCTG ATGTTTGATG CCGGCACGGT TGCCGCCTTG

5

10

15

This corresponds to the amino acid sequence <SEQ ID 486; ORF53ng-1>:

20

25

ORF53ng-1 and ORF53-1 show 94.0% identity in 336 aa overlap:

3

3

4

4

4

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 58

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 487>:

```

1  ..TTGCGGGAAA CGGCATATGT TTTGGATAGT TTTGATCGTT ATTTTGTGT
51  TGCGCTTGCC GGCTTGT TTTT TGTCCGCGC ACAATCCGAA CGCGAGTGGA
101 TGCGCGAGGT TTCTGCGTGG CAGGAAAAGA AAGGGGAAAA ACAGGCGGAG
151 CTGCCTGAAA TCAAAGACGG TATGCCCGAT TTTCCCGAAC TTGCCCTGAT
201 GCTTTTCCAC GCCGTCAAAA CGGCAGTGT TTTGGCTGTT GTCCGTGTCG
251 TCCGTTTCTG CCGAACTAT CTGGCGCAGC AATCCGAACC GGACAGGCC
301 GTCCGCCT..

```

This corresponds to the amino acid sequence <SEQ ID 488; ORF58>:

```

1  ..LRETAYVLD FDRYFVVALA GLFFVRAQSE REWMREVS AW QEKKGEKQAE
51  LPEIKDGMFD FPELALMLEH AVKTAVYWLF VGVVRF CRNY LAHESEPD RP
101  VPP..

```

Further work revealed the complete nucleotide sequence <SEQ ID 489>:

```

20  1  ATGTTTGGGA TAGTTTGGAT CGTTATTTTG TTGCTTGC GC TTGCCGGCTT
51  GTTTTTGTGC CGCGCACAA TCCGAACGCGA GTGGATGCGC GAGGTTTCTG
101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA
151 GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT
201 CAAAACGGCA GTGTATTGGC TGTTTGTCCG TGTGCTCCGT TTCTGCCGAA
25  ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
301 GCAAACCGTG CGGATTTTCC GACCGCATCC GACGGATATT CAGACAGTGG
351 AAACGGGACG GAAGAAGCGG AAACGGAAGA AGCAGAAGCT GCGGAGGAAG
401 AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
451 ATCCCATTCG ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
30  501 AATTTGCCCC GTCCGTCCGG TTTTAAAGA AATCACTTTG GAAGAAGCAA
551 CGCGTGCTTT AAACAGCGCG GCTTTAAGGG AAACGAAAAA ACGCTATATC
601 GATGCATTTC AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCCGA
651 TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
701 AACGCACGTA TTCCCATATG TTCGATGCGG ACAAGAAGC GTTTTCCGAG
35  751 TCTGCGGATT ACGGATTTGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC
801 CTTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCCACGTC
851 ATGCAGGGCA GGGGAAAGGG CAGGCGGAGG CAAAATCCCC GGATGTTTCC
901 CAAGGGCAGT CCGTTTCAGA CGGCACGGCC GTCCGCGATG CCCGCCGCCG
951 CGTTTCCGTC AATTTGAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG
40  1001 CGCGAATTTC TCGCCTGATT CCGGAAAGTC AGACGGTTGT CGGGAACCGG
1051 GATGTCGAAA TGCCGTCTGA AACCGAAAAT GTTTTACGCG AAACCGTTTC
1101 GTCTGTGGGA TACGGCGGTC CGGTTTATGA TGAACTGCC GATATCCATA
1151 TTGAAGAACC TGCCGCGCCC GATGCTTGGG TGGTCGAACC ACCCGAAGTG
1201 CCGAAAGTTC CCATGACCGC AATCGATATT CAGCCGCCGC CTCCCGTATC
45  1251 GGAAATCTAC AACCGTACCT ATGAACCGCC GTCAGGATTC GAGCAGGTGC
1301 AACGCAGCCG CATTGCCGAG ACCGACCATC TTGCCGATGA TGTTTTGAAT
1351 GGAGGTTGGC AGGAGGAAAC CGCCGCTATT GCGGATGACG GCAGTGAAGG
1401 TGCGGCAGAG CGGTCAAGCG GGCAATATCT GTCGGAACC GAAGCGTTCG
1451 GGCGATGACG TCAGGCGGTT TGTCCGTTG AAAATGTGCC GTCTGAACGC
50  1501 CCGTCCTGCC GGTATCCGA TACGGAAGCG GATGAAGGGG CGTTCCCATC
1551 TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
1601 TGCCTCCGCT GTTCAATCCC GAGGCGACGC AAACCGAAGA AGAACTGTTG
1651 GAAAACAGCA TCACCATCGA AGAAAAATTG GCGGAGTTCA AAGTCAAGGT
1701 CAAGGTTGTC GATTCTTATT CCGGCCCGGT AATTACGCGT TATGAAATCG
55  1751 AACCCGATGT CGGCGTGC GC CAATCCG TTCTGAATCT GGAAAAAGAT
1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGCGTTGTCG AAACCATCCC
1851 CGGCAAAACC TGCGTGGGTT TGGAACTTCC GAACCGGAAA CGCCAAATGA

```

-287-

1901 TACGCCTGAG CGAAATCTTC AATTCGCCCC AGTTTGCCGA ATCCAAATCC
 1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCGTAAAC
 2001 CGACTTGGGA AAAGCACC GC ATTTGTTGGT TGCCGGCAGC ACCGGTTCGG
 2051 GCAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTCAAAGCC
 5 2101 GCGCCGGAAG ACGTGCCTAT GATTATGATC GATCCGAAAA TGCTGGAATT
 2151 GAGCATTAC GAAGGCATCC CGCACCTGCT CGCCCCTGTC GTTACCGATA
 2201 TGAAGCTGGC GGCAACGCG CTGAACTGGT GTGTTAACGA AATGGAAAAA
 2251 CGCTACCGCC TGATGAGCTT TATGGGCGTG CGTAATCTTG CCGGCTTCAA
 10 2301 TCAAAAAATC GCCGAAGCCG CAGCAAGGGG AGAAAAATC GGCAATCCGT
 2351 TCAGCTCAC GCCCGACGAT CCCGAACCTT TGGAAAACT GCCGTTTATC
 2401 GTGGTCGTGG TCGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA
 2451 AATCGAAGAA CTGATTGCCC GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA
 2501 TCCATTGAT TCTTGCCACA CAACGCCCA GCGTCGATGT CATCACGGGT
 2551 CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA
 15 2601 AATCGACAGC CGCAGGATTC TCGACCAAT GGGCGCGGAA AACCTGCTCG
 2651 GTCAGGGCGA TATGCTGTTT CTGCTGCCGG GTACTGCCTA TCCGACGCGC
 2701 GTTACGCGC CGTTTGCCTC GGATGAAGAG GTGCACCGCG TGGTCGAATA
 2751 TTTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATT TTGAGCGGCG
 2801 GCGGCAGCGA AGAGCTGCCC GGCATCGGGC GCAGCGGCGA CGACGAAACC
 2851 GATCCGATGT ACGACGAGGC CGTATCCGTT GTCCTGAAAA CGCGCAAAGC
 20 2901 CAGCATTTCG GGCCTACAGC GCGCCTTGGC TATCGGCTAC AACC CGCCG
 2951 CGCGTCTGAT TGACCAGATG GAGCGGAAG GCATTGTGTC CGCACCGGAA
 3001 CACAACGCA ACCGTACGAT TCTCGTCCCC TTGACAATG CTGTA

This corresponds to the amino acid sequence <SEQ ID 490; ORF58-1>:

25 1 MFIVILVIL LLALAGLFFV RAQSEREWMR EVSAWQEKKG EKQALPEIK
 51 DGMDFPELA LMLFHAVKTA VYWLFGVVR FCRNYLAHES EPDRPVPPAS
 101 ANRADVPTAS DGYSDSGNGT EEAETEEAEA AEEEEADTED IATAVIDNRR
 151 IPFDRSIAEG LMPSESEISP VRPVFKEITL EATRNLNSA ALRETKKRYI
 201 DAFEKNETAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSHM FDADKEAFSE
 30 251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FHRHAGQKGQ QAEAKSPDVS
 301 QGQVSDGTA VRDARRRVSV NLKEPNKATV SAEARISRLI PESQTVVGKR
 351 DVEMPSETEN VFTETVSSVG YGGPVYDETA DIHIEEPAP DAWVEPPEV
 401 PKVPMTAIDI QPPPPVSEIY NRTYEPSPGF EQVQRSRIAE TDHLADDVLN
 451 GGWQEETAAI ADDGSEGAAE RSSGOYLSET EAFGHDSQAV CPFENVPSER
 501 PSCRVSDETA DEGAFFSEET GAVSEHLPTT DLLLPPLFNP EATQTEEELL
 35 551 ENSITIEEKL AEFKVKVKV DSYSGPVITR YEIEPDVGVR GNSVLNLEKD
 601 LARSLGVASI RVVETIPGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS
 651 KLTALGQDI TGQPVVTDLG KAPHLLVAGT TSGSKSVGVN AMILSMLFKA
 701 APEDVRMIMI DPKMLELSIY EGIPHLLAPV VTDMKLAANA LNWCVNEMEK
 40 751 RYRLMSFMGV RNLAGFNQKI AEAARGEKI GNPFSLTPDD PEPLEKLPEI
 801 VVVDEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPSVDVITG
 851 LIKANIPTRI AFQVSSKIDS RTILDQMGAE NLLGQDMLF LLPGTAYPQR
 901 VHGAFADEE VHRVVEYLKQ FGEPDYVDDI LSGGGSEELP GIGRSGDDET
 951 DPMYDEAVSV VLKTRKASIS GVQRLRIGY NRAARLIDQM EAEGIVSAPE
 45 1001 HNGNRTILVP LDNA*

Computer analysis of this amino acid sequence predicts the indicated transmembrane region, and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF58 shows 96.6% identity over a 89aa overlap with an ORF (ORF58a) from strain A of *N.*

50 *meningitidis*:

10 20 30 40 50 60
 orf58.pep LRETAYVLDSFDTRYFVVALAGLFFVRAQSEREWMREVS AWQEKKG EKQALPEIKDGM
 orf58a MFIVILVILLLALAGLFFVRAQSEREWMREVS AWQEKKG EKQALPEIKDGM
 55 10 20 30 40 50
 orf58.pep FPELALMLFHAVKTAVYWLFGVVRFCRNYLAHESEPD RPVP
 orf58a FPELALMLFHAVKTAVYWLFGVVRFCRNYLAHESEPD RPVP PASANRADVPTASDGYSD

The complete length ORF58a nucleotide sequence <SEQ ID 491> is:

```

1  ATGTTTGGGA TAGTTTGGAT CGTTATTTTG TTGCTTGCGC TTGCCGGCTT
51  GTTTTTTGTC CGCGCACAAAT CCGAACGCCA GTGGATGCGC GAGGTTTCTG
101 CGTGGCAGGA AAAGAAAGGG GAAAAACAGG CGGAGCTGCC TGAAATCAAA
151 GACGGTATGC CCGATTTTCC CGAACTTGCC CTGATGCTTT TCCATGCCGT
201 CAAAACGGCA GTGTATTGGC TGTGTGTCGG TGTCGTCCGT TTCTGCCGAA
251 ACTATCTGGC GCACGAATCC GAACCGGACA GGCCCGTTCC GCCTGCTTCT
301 GCAAATCGTG CGGATGTTCC GACCGCATCC GACGGATATT CAGACAGTGG
351 AAACGGGACG GAAGAAGCGG AAACGGGAAG AGCAGAAGCT GCGGAGGAAG
401 AGGCTGCCGA TACGGAAGAC ATTGCAACTG CCGTAATCGA CAACCGCCGC
451 ATCCCATTCG ACCGGAGTAT TGCTGAAGGG TTGATGCCGT CTGAAAGCGA
501 AATTTCCGCC GTCCGTCCGG TTTTAAAGGA AATCACTTG GAAGAAGCAA
551 CGCGTGCTTT AAACAGCGCG GCTTAAAGGG AAACGAAAAA ACGCTATATC
601 GATGCATTTG AGAAAAACGA AACAGCGGTC CCCAAAGTCC GCGTGTCCGA
651 TACCCCGATG GAAGGGCTGC AGATTATCGG TTTGGACGAC CCTGTGCTTC
701 AACGCACGTA TTCCCGTATG TTCGATGCGG ACAAAGAAGC GTTTCCGAG
751 TCTGCGGATT ACGGATTTGA GCCGTATTTT GAGAAGCAGC ATCCGTCTGC
801 CTTTTCTGCA GTCAAAGCCG AAAATGCACG GAATGCGCCG TTCGCCGCTC
851 ATGCAGGGCA GGGNAAAGGG CAGGCGGAGG CNAATCCCC GGATGTTTCC
901 CAAGGGCAGT CCGTTTCAGA CGGCACAGCC GTCCGCGATG CCNGCCGCCG
951 CGTTTCCGTC AATTTGAAAG AACCGAACAA GGCAACGGTT TCTGCGGAGG
1001 GCGGGATTTC GCGCCTGATT CCGGAAAGTC GGACGGTTGT CGGGAACCGG
1051 GATGTCGAAA TGCCGTCTGA AACCGAAAAT GTTTTCACGG AAANTGTTTC
1101 GTCTGTGGGA TACGGCGNTC CGGTTTATGA TGAACTGCC GATATCCATA
1151 TTGAAGAACG TGCCGCGCCC GATGCTTGGG TGTCGAACC ACCCGAAGTG
1201 CGGAAAGTTC CCATGCCCGC AATNGATATT CCGCCGCCGC CTCCCGTATC
1251 GGAAATCTAC AACCGTACCT ATGAACCGCC GGCAGGATTC GAGCAGGTGC
1301 AACGCACCGC CATTGCCGAA ACCGATCATC TTGCCGATGA TGTTTGAAT
1351 GGAGTTGGC AGGAGGAAAC CGCCGCTATT GCGAATGACG GCAGTGAGGG
1401 TGTGGCAGAG CGGTCAAGCG GGCAATATT GTCGAAACC GAAGCGTTCG
1451 GGCATGACAG TCAGGCGGTT TGTCCGTTG AAAATGTGCC GTCTGAACGC
1501 CCGTCCCGCC GGGCATNGGA TACGGAAGCG GATGAAGGGG CGTCCAATC
1551 TGAAGAAACC GGTGCGGTAT CCGAACACCT GCCGACAACC GACCTGCTTC
1601 TGCCGCGGCT GTTCAATCCC GGGCGACGC AAACCGAAGA AGANCTGTTG
1651 GANAACAGCA TCACCATCGA AGAAAAATNG GCGGAGTTCA AAGTCAAGGT
1701 CAAGGTTGTC GATTCTTATT CCGGCCCGCT GATTACCGT TATGAAATCG
1751 AACCCGATGT CGGCGTCCGC GGCAATTCCG TTCTAAATCT GGAAAAAGAN
1801 TTGGCGCGTT CGCTCGGCGT GGCTTCCATC CGGTTGTGCG AAACCATCTC
1851 CGGCAAACC TGTATGGGTT TGGAACTTCC GAACCCGAAA CGCCAAATGA
1901 TACGCCTGAG CGAAATCTTC AATTCGCCCG AGTTTGCCGA ATCCAAATCC
1951 AAGCTGACGC TCGCGCTCGG TCAGGACATC ACCGGACAGC CCGTCGTAAC
2001 CGACTTGGGC AAAGCACCGC ATTTGTTGGT TGCCGGCAGC ACCGTTTCGG
2051 GCAAATCGGT GGGTGTCAAC GCGATGATTC TGTCTATGCT TTCAAAGCC
2101 GCGCCGGAAG ACGTGCGTAT GATTATGATC GATCCGAAA TGCTGGAATT
2151 GAGCATTTAC GAAGGCATCC CGCACCTGCT CGCCCTGTC GTTACCGATA
2201 TGAAGCTGGC GGCAAACGCG CTGAACTGGT GTGTTAACGA AATGGAAAAA
2251 CGCTACGCC TGATGAGCTT TATGGGCGTG CGCAATCTG CCGGTNTCAA
2301 TCAAAAAATC GCCGAAGCCG CAGCAAGGGG GGAGAAAATC GGCAACCCGT
2351 TCAGCCTCAC GCCCGACAAT CCGGAACCTT TGGANAATT GCCGTTTATC
2401 GTGGTCTGG TTGATGAGTT TGCCGACCTG ATGATGACGG CAGGCAAGAA
2451 AATCGAAGAA CTGATTGCCG GCCTCGCCCA AAAAGCCCGC GCGGCAGGCA
2501 TCCATCTTAT CCTTGCCACA CAACGCCCCA GTGTCGATGT CATCACGGGT
2551 CTGATTAAGG CGAACATCCC GACGCGTATC GCGTTCCAAG TGTCCAGCAA
2601 AATCGACAGC CGCACGATTC TTGACCAAT GGGTGGCGAA AACCTGCTCG
2651 GGCAGGGCGA TATGCTGTTT CTGCCGCCGG GTACGGCTA TCCGCAGCCG
2701 GTTCACGGCG CGTTTGCCCTC GGATGAAGAG GTGCACCGCG TGGTGAATA
2751 TCTGAAACAG TTTGGCGAAC CGGACTATGT TGACGATATN TTGAGCGGCG
2801 GTATGTCCGA CGATTTGCTG GGAATCAGCC GGAGCGGCGA CCGCGAAACC
2851 GATCCGATGT ACGACGAGGC CGTGTCTGTT GTTTGAAAA CGCGCAAAGC
2901 CAGCATTTCT GCGGTGCAGC GCGCATTGCG TATCGGCTAT AATCGCGCGC
2951 CGCGTCTGAT TGACCAGATG GAGGCGGAAG GCATTGTGTC CGCACCGGAA
3001 CACAACGGCA ACCGTACGAT TCTCGTCCCC TTNGACAATG CTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 492>:

```

1  MFIVLVIL LLAGLFFV RAQSEREWMR EVSAWQEKKG EKQALPEIK
65  DGMPDFPELA LMLFHAVKTA VYWLFGVVR FCRNYLAHES EPDRPVPPAS
101 ANRADVPAS DGYSDSNGT EEAETEEAEA AEEEAADTED IATAVIDNRR
151 IPFDRSIAEG LMPSESEISP VRPVFKEITL EEATRALNSA ALRETKKRYI
201 DAFEKNETAV PKVRVSDTPM EGLQIIGLDD PVLQRTYSRM FDADKEAFSE

```

5 251 SADYGFEPYF EKQHPSAFSA VKAENARNAP FRRHAGQGKG QAEAKSPDVS
 301 QGQSVSDGTA VRDAXRRVSV NLKEPNKATV SAEARISRLI PESRTVVGKR
 351 DVEMPSETEN VFTEXVSSVG YGXPVYDETA DIHIEPAAP WDAWVVEPPEV
 401 PKVPMPAXDI PPPPVSEIY NRTYEPPAGF EQQRSRIAE TDHLADDVLN
 451 GGWQEETAAI ANDGSEGVAE RSSGOYLSET EAFGHDSQAV CPFENVPSER
 501 PSRRAXDTEA DEGAFAQSEET GAVSEHLPTT DLLLPPLFNP GATQTEEXLL
 551 XNSITIEEKX AEFKVKVKVV DSYSGPVITR YEIEPDVGVR GNSVLNLEKX
 601 LARSLGVASI RVVETILGKT CMGLELPNPK RQMIRLSEIF NSPEFAESKS
 10 651 KLTLALGQDI TGQPVVTDLG KAPHLLVAGT TGSQKSVGVN AMILSMLFKA
 701 APEDVRMIMI DPKMLELSIY EGIPHLLAPV VTDMLAANA LNWCVNEMEK
 751 RYRLMSFMGV RNLAGXNQKI AEAARGEKI GNPFSLTPDN PEPLXKLPI
 801 VVVVDEFADL MMTAGKKIEE LIARLAQKAR AAGIHLILAT QRPSVDVITG
 851 LIKANIPTRI AFQVSSKIDS RTILDOMGAE NLLGQGDMLF LPPGTAYPQR
 901 VHGAFADEE VHRVVEYLKQ FGEPDYVDDX LSGGMSDDL GISRSGDGET
 15 951 DPMYDEAVSV VLKTRKASIS GVQRALRIGY NRAARLIDQM EAEGIVSAPE
 1001 HNGNRTILVP XDNA*

ORF58a and ORF58-1 show 96.6% identity in 1014 aa overlap:

		10	20	30	40	50	60
20	orf58a.pep	MFWIVLIVILLLLALAGLFFVRAQSEREWREVSAAWQEKKGKQAEKPEIKDGMDFPELA					
	orf58-1	MFWIVLIVILLLLALAGLFFVRAQSEREWREVSAAWQEKKGKQAEKPEIKDGMDFPELA					
		10	20	30	40	50	60
25	orf58a.pep	LMLFHAVKTAVYWLFGVVRFCRNYLAHESEPDPRVPPASANRADVPTASDGYSDSGNGT					
	orf58-1	LMLFHAVKTAVYWLFGVVRFCRNYLAHESEPDPRVPPASANRADVPTASDGYSDSGNGT					
		70	80	90	100	110	120
30	orf58a.pep	EEAEETEEAEAEAAADTEDIATAVIDNRRIPFDRSIAEGLMPSESEISPRVPVFKEITL					
	orf58-1	EEAEETEEAEAEAAADTEDIATAVIDNRRIPFDRSIAEGLMPSESEISPRVPVFKEITL					
		130	140	150	160	170	180
35	orf58a.pep	EEATRALNSAALRETKKRYIDAFKNETAVPKVRVSDTPMEGLQIIGLDDPVLQRTYSRM					
	orf58-1	EEATRALNSAALRETKKRYIDAFKNETAVPKVRVSDTPMEGLQIIGLDDPVLQRTYSRM					
		190	200	210	220	230	240
40	orf58a.pep	FDADKEAFSESADYGFEPYFEKQHPSAFSAVKAENARNAPFRRHAGQGKGQAEAKSPDVS					
	orf58-1	FDADKEAFSESADYGFEPYFEKQHPSAFSAVKAENARNAPFRRHAGQGKGQAEAKSPDVS					
		250	260	270	280	290	300
45	orf58a.pep	QGQSVSDGTAVRDAXRRVSVNLKEPNKATVSAEARISRLIPESRTVVGKRDVEMPSETEN					
	orf58-1	QGQSVSDGTAVRDAXRRVSVNLKEPNKATVSAEARISRLIPESRTVVGKRDVEMPSETEN					
		310	320	330	340	350	360
50	orf58a.pep	VFTEXVSSVGYGXPVYDETA DIHIEPAAPDAWVVEPPEV PKVPMPAXDIPPPPVSEIY					
	orf58-1	VFTEVSSVGYGXPVYDETA DIHIEPAAPDAWVVEPPEV PKVPMTAIDIQPPPVSEIY					
		370	380	390	400	410	420
55	orf58a.pep	NRTYEPPAGFEQVQRSRIAE TDHLADDVLNGGWQEETAAIANDGSEGVAE RSSGOYLSET					
	orf58-1	NRTYEPPSGFEQVQRSRIAE TDHLADDVLNGGWQEETAAIADDGSEGAERSSGOYLSET					
		430	440	450	460	470	480
60	orf58a.pep	EAFGHDSQAVCPFENVPSERPSRRAXDTEADGAFQSEETGAVSEHLPTT DLLLPPLFNP					
	orf58-1	EAFGHDSQAVCPFENVPSERPSRVSDTEADGAFSEETGAVSEHLPTT DLLLPPLFNP					
		490	500	510	520	530	540
65	orf58a.pep	EAFGHDSQAVCPFENVPSERPSRVSDTEADGAFSEETGAVSEHLPTT DLLLPPLFNP					
	orf58-1	EAFGHDSQAVCPFENVPSERPSRVSDTEADGAFSEETGAVSEHLPTT DLLLPPLFNP					
		490	500	510	520	530	540
70	orf58a.pep	EAFGHDSQAVCPFENVPSERPSRVSDTEADGAFSEETGAVSEHLPTT DLLLPPLFNP					
	orf58-1	EAFGHDSQAVCPFENVPSERPSRVSDTEADGAFSEETGAVSEHLPTT DLLLPPLFNP					

5	orf58a.pep	550 560 570 580 590 600	GATQTEEXLLXNSITIEEKXAEFKVKVVDYSYSGPVITRYEIEPDVGVRGNSVLNLEKX
	orf58-1	550 560 570 580 590 600	EATQTEEELENSITIEEKLAEFKVKVVDYSYSGPVITRYEIEPDVGVRGNSVLNLEKD
10	orf58a.pep	610 620 630 640 650 660	LARSLGVASIRVVETILGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDI
	orf58-1	610 620 630 640 650 660	LARSLGVASIRVVETIPGKTCMGLELPNPKRQMIRLSEIFNSPEFAESKSKLTALGQDI
15	orf58a.pep	670 680 690 700 710 720	TGQPVVTDLGKAPHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY
	orf58-1	670 680 690 700 710 720	TGQPVVTDLGKAPHLLVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIY
20	orf58a.pep	730 740 750 760 770 780	EGIPHLLAPVVTDMLAANALNWCVNEMEKRYRLSMFGVRNLAGXNQKIAEAAARGEKI
	orf58-1	730 740 750 760 770 780	EGIPHLLAPVVTDMLAANALNWCVNEMEKRYRLSMFGVRNLAGFNQKIAEAAARGEKI
25	orf58a.pep	790 800 810 820 830 840	GNPFSLTPDNPEPLXKLPIVVVVDEFADLMTAGKKIEELIARLAQKARAAGIHLILAT
	orf58-1	790 800 810 820 830 840	GNPFSLTPDDPEPLEKLPIVVVVDEFADLMTAGKKIEELIARLAQKARAAGIHLILAT
30	orf58a.pep	850 860 870 880 890 900	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLPPGTAYPQR
	orf58-1	850 860 870 880 890 900	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLPPGTAYPQR
35	orf58a.pep	910 920 930 940 950 960	VHGAFASDEEVHRVVEYLKQFGEPTYVDDXLSSGMSDDLIGISRGSDGETDPMYDEAVSV
	orf58-1	910 920 930 940 950 960	VHGAFASDEEVHRVVEYLKQFGEPTYVDDILSGGSEELPGIGRSGDDETDPYDEAVSV
40	orf58a.pep	970 980 990 1000 1010	VLKTRKASISGVQRALRIGYNRAARLIDQMEAGIVSAPEHNGNRTILVLPXDNAX
	orf58-1	970 980 990 1000 1010	VLKTRKASISGVQRALRIGYNRAARLIDQMEAGIVSAPEHNGNRTILVPLDNAX

50 Homology with a predicted ORF from *N.gonorrhoeae*

ORF58 shows complete identity over a 9aa overlap with a predicted ORF (ORF58ng) from *N. gonorrhoeae*:

55	orf58.pep	ALMLFHAVKTAVYWLFGVVRFCRNYLAHESEPD RPVP	103
	orf58ng	SEPDRPVPPASANRADVPTASDGYSDSGNG	30

The ORF58ng nucleotide sequence <SEQ ID 493> is predicted to encode a protein having partial amino acid sequence <SEQ ID 494>:

60	1	SEPDRPVPPA SANRADVPTA SDGYSDSGNG TEEAEETAAE AAEEEEAADTE
	51	DIATAVIDNR RIFDRSIAE GLMQSESKTS PVRPVFKEIT LEEATRALSS
	101	AALRETKKRY IDAFEKNGTA VPKVRVSDTP MEGLOIIGLO DPVLQRTYSR
	151	MFDADKEAFS ESADYGFEFY FEKQHPSAFS AVKAENARNA PFRRHAGQEK
	201	GQAEAKSPDV SQGQSVSDGT AVRDRARRVS VNLKEPNKAT VSAEARISRL
	251	IPESRTVVGK RDVEMPSET E NVFTETVSSV GYGGPVYDEA ADIHIEEPAA
	301	PDWVVEPPE VFEVAVPEID ILPPPPVSEI YNRTYEPPAG FEQAQRSRIA

351	ETDHLAADVL	NGGWQEETAA	IADGSEGA	ERSGQYLSE	TEAFGHDSQA
401	VCPFEDVPSE	RPSCRVSDE	ADEGAFQSEE	TGAVSEHLPT	TDLLLPPLFN
451	PEATQTEEL	LENSITIEEK	LAEFKVKVKV	VDSYSGPVIT	RYEIEPDVGV
501	RGNSVLNLEK	DLARSLGVAS	IRVVETIPGK	TCMGLELPNP	KRQMIRLSEI
551	FNSPEFAESK	SKLTLAGQD	ITGQPVVTDL	GKAPHLVAG	TTGSGKSVGV
601	NAMILSMLFK	AAPEDVRMIM	IDPKMLELSI	YEGITHLLAP	VVTDMKLAAN
651	ALNWCVNEME	KRYRLMSFMG	VRNLAGFNQK	IAEAAARGEK	IGNPFSLTDP
701	DPEPLEKLFF	IVVVVDEFAD	LMMTAGKKIE	ELIARLAQKA	RAAGIHLILA
751	TQRPSVDVIT	GLIKANIPTR	IAFQVSSKID	SRTILDQMG	ENLLGQDML
801	FLPPGTAYPQ	RVHGAFASDE	EVHRVVEYLK	QFGEPTYDD	ILSGGSGEEL
851	PGIGRSGDGE	TDPMYEAVS	VVLKTRKASI	SGVQRALRIG	YNRAARLIDQ
901	MEAEIGVSAP	EHNGNRTILV	PLDNA*		

This partial gonococcal sequence contains a predicted transmembrane region and a predicted ATP/GTP-binding site motif A (P-loop; double underlined). Furthermore, it has a domain homologous to the FTSK cell division protein of *E. coli*. Alignment of ORF58ng and FtsK (accession number p46889) show a 65 % amino acid identity in 459 overlap:

ORF58ng:	467	IEEKLAEFKVKVVD	SYSGPVITRYEIEPDV	GVGRNSVLNLEKDLARSLG	VASIRVVET	526
		+E +LA+F++K	VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE			
FtsK:	868	VEARLADFRKADV	VNYSPPGVITRFE	LNLAGPVKAA	RISNLSRDLARSL	STVAVRVVEV 927
ORF58ng:	527	IPGKTCMGLELPN	KRQMIRLSEIFNSPEFAESK	SKLTLAGQDITGQPVVTD	LGKAPHL	586
		IPGK +GLELPN	KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL			
FtsK:	928	IPGKPYVGLELPN	KKRQTVYLREVL	NAKFRDNPSPLTVVLG	KDIAEPV	VADLAKMPHL 987
ORF58ng:	587	LVAGTTGSGKSVG	VNAMILSMLFKAAPEDVR	MIMIDPKMLELSIYEGITHLLAP	VPVTD	DMK 646
		LVAGTTGSGKSVG	VNAMILSML+KA PEDVR	IMIDPKMLELS+YEGI HLL	VVTDMK	
FtsK:	988	LVAGTTGSGKSVG	VNAMILSMLYKAQPEDVR	FIMIDPKMLELSVYEGIPHL	LTVEV	TDK 1047
ORF58ng:	647	LAANALNWCVNEME	KRYRLMSFMGVRNLAGFNQKIAEAAARGEK	IGNPFSLT	PDDPEP--	704
		AANAL WCVNEME+RY+LMS +GVRNLAG+N+KIAEA	I +P+ D +			
FtsK:	1048	DAANALRWCVNEM	ERRYKLMSALGVRNLAGYNEKIAEADRM	MRPIPDPYWKPGDSMDAQH		1107
ORF58ng:	705	--LEKLPFIVVV	DEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT	QRP	SDVITGL	762
		L+K P+IVV+VDEFADLMMT	GKK+EEELIARLAQKARAAGIHL+LAT	QRP	SDVITGL	
FtsK:	1108	PVLKKEPYIVVL	VDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLAT	QRP	SDVITGL	1167
ORF58ng:	763	IKANIPTRIAFQV	SSKIDSRTILDQMAENLLGQDMLFLPPGTAYPQ	RVHGAFASDEEV		822
		IKANIPTRIAF VSSKIDSRTILDQ	GAE+LLG GDML+ P + P RVHGAF D+EV			
FtsK:	1168	IKANIPTRIAFTV	SSKIDSRTILDQAGAESLLGMGDMLYSGPNSTLP	VRVHGAFVRDQEV		1227
ORF58ng:	823	HRVVEYLKQFGE	PDYVDDILSGGSGEELPGIGRSGDGETDPMYDEAVS	VVLKTRKASISG		882
		H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG				
FtsK:	1228	HAVVQDWKARGRP	QYVDGITSDESEGGAG-GFDGAEELDPLFDQAVQFVTEKR	KASISG		1286
ORF58ng:	883	VQRALRIGYNRAAR	LIDQMEAEIGVSAPEHNGNRTILVP			921
		VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P				
FtsK:	1287	VQRQFRIGYNRAARI	EQMEAQGVSEQGHNGNREVLAP			1325

Further work on ORF58ng revealed the complete gonococcal DNA sequence to be <SEQ ID 495>:

1	ATGTTTGGGA	TAGTTTGGAT	CGTTATgtg	TTGCTTGCGC	TTGCCGCGCT
51	GTTTTTGTG	CGCGACAAT	CCGAACGCGA	GTGGATGCGC	GAGGTTTCTG
101	CGTGGCAGGA	AAAGAAAGGG	GAAAAACAGG	CGGAGCTGCC	TGAAATCAAA
151	GACGGTATGC	CCGATTTTCC	CGAGTTTTC	CTGATGCTTT	TCCATGCCGT
201	CAAAACGGCA	GTGTATTGGC	TGTTTGTCGG	TGTCGTCCGT	TTCTGCCGAA
251	ACTATCTGGC	GCACGAATCC	GAACCGGACA	GGCCCGTTCC	GCCTGCTTCT
301	GCAAACCGTG	CGGATGTTCC	GACCGCATCC	GACGGGTATT	CAGACAGTGG
351	AAACGGGACG	GAAGAAGCGG	AAACGGAAGC	AGCAGAAGCT	GCGGAGGAAG
401	AGGCTGCCgA	TACgGAAGAC	ATTGCAACTG	CCGTAATCGA	CAACCGCCGC
451	ATCCcatTCG	ACCGGAGTAT	TGCTGAAGGG	TTGATGCACT	CTGAAAGCAA
501	AACTTCGCCC	GTCCGTCCGG	TTTTTAAGGA	AATCACTTTG	GAAGAAGCAA
551	CGCGTGCTTT	AAGCAGCGCG	GCTTTAAGGG	AAACGAAAAA	ACGCTATATC
601	GATGCATTTG	AGAAAAACGG	AACAGCCGTC	CCCAAAGTAC	GCGTGTCCGA
651	TACCCCGATG	GAAGGGCTGC	AGATTATCGG	TTTGACGAC	CCTGTGCTTC
701	AACGCACGTA	TTCCCGTATG	TTTGATGCGG	ACAAAGAAGC	GTTTTCCGAG
751	TCTGCGGATT	ACGGATTGA	GCCGTATTTT	GAGAAGCAGC	ATCCGTCTGC

801	CTTTTCTGCA	GTCAAAGCCG	AAAATGCACG	GAATGCGCCG	TTCCGCCGTC
851	ATGCAGGGCA	GGAGAAAGGG	CAGGCGGAGG	CAAAATCCCC	GGATGTTTCC
901	CAAGGGCAGT	CCGTTTCAGA	CGGCACAGCC	GTCCGCGATG	CCCGCCGCCG
951	CGTTTCCGTC	AATTTGAAAG	AACCGAACAA	GGCAACGGTT	TCTGCGGAGG
1001	CGCGGATTTC	GCGCCTGATT	CCGGAAGATC	GGACGTTTGT	CGGGAACCGG
1051	GATGTCGAAA	TGCCGTCTGA	AACCGAAAAT	GTTTTCACGG	AAACCGTTTC
1101	GTCTGTGGGA	TACGGCGGTC	CGGTTTATGA	TGAAGCTGCC	GATATCCATA
1151	TTGAAGAGCC	TGCCGCGCCC	GATGCTTGGG	TGGTCGAACC	ACCCGAAGTG
1201	CCGGAGGTAG	CCGTACCCGA	AATCGATATT	CTGCCGCCGC	CTCCCGTATC
1251	GGAAATCTAC	AACCGTACCT	ATGAGCCGCC	GGCAGGATTC	GAGCAGGCGC
1301	AACGCAGCCG	CATTGCCGAA	ACCGACCATC	TTGCCGCTGA	TGTTTGAAT
1351	GGAGGTTGGC	AGGAGGAAAC	CGCCGCTATT	GCAGATGACG	GCAGTGAGGG
1401	TGCGGCAGAG	CGGTCAAGCG	GGCAATATCT	GTCGGAACCC	GAAGCGTTCC
1451	GGCATGACAG	TCAGGCGGTT	TGTCCGTTTG	AAGATGTGCC	GTCTGAACGC
1501	CCGTCTCGCC	GGGTATCGGA	TACGGAAGCG	GATGAAGGGG	CGTTCCAATC
1551	GGAAGAGACC	GGTGCCGTAT	CCGAACACCT	GCCGACAACC	GACCTGCTTC
1601	TGCCTCCGCT	GTTCAATCCC	GAGGCGACGC	AAACCGAAGA	AGAACTGTTG
1651	GAAAAACAGCA	TCACCATCGA	AGAAAAATTG	GCGGAGTTCA	AAGTCAAGGT
1701	CAAGGTTGTC	GATTCTTATT	CCGCCCCCGT	GATTACGCGT	TATGAAATCG
1751	AACCCGATGT	CGGCGTGCGC	GGCAATTCCG	TTCTGAATTT	GGAAAAAGAC
1801	TTGGCGCGTT	CGCTCGGCGT	GGCTTCCATC	CGCGTTGTTC	AAACCATCCC
1851	CGGCAAAACC	TGCATGGGTT	TGGAACCTCC	GAACCCGAAA	CGCCAAATGA
1901	TACGCTGAG	CGAAATTTTC	AATTGCGCCG	AGTTTGCCGA	ATCCAAATCC
1951	AAGCTGACGC	TCGCGCTCGG	TCAGGACATT	ACCGGACAGC	CCGTCGTAAC
2001	CGACTTGGGC	AAAGCACCCG	ATTGCTGGT	TGCCGGCAGC	ACCGGTTCCG
2051	GCAAATCCGT	GGGTGTCAAC	GCGATGATTC	TGCTATGCT	TTTCAAAGCC
2101	GCGCCGGAAG	ACGTGCGTAT	GATTATGATC	GATCCGAAAA	TGCTGGAATT
2151	GAGCATTTAC	GAAGGCATCA	CGCACCTGCT	CGCCCTGTCT	GTTACCGATA
2201	TGAAGCTGGC	GGCAAAACGC	CTGAACCTGG	GTGTTAACGA	AATGGAAAAA
2251	CGCTACCGCC	TGATGAGCTT	TATGGGCGTG	CGCAATCTTG	CGGGCTTCAA
2301	CCAAAAAATC	GCCGAAGCCG	CAGCAAGGGG	AGAAAAAATC	GGCAATCCGT
2351	TCAGCCTCAC	GCCCGACGAT	CCCGAACCTT	TGGAAAAACT	GCCGTTTATC
2401	GTGGTCGTGG	TCGATGAGTT	TGCCGATTTC	ATGATGACGG	CAGGCAAGAA
2451	AATCGAAGAA	CTGATTGCGC	GCCTCGCCCA	AAAAGCCCGC	GCGGCAGGCA
2501	TCCACCTTAT	CCTTGCCACA	CAACGCCCCA	GCGTCGATGT	CATCACGGGT
2551	CTGATTAAGG	CGAACATCCC	GACGCGTATC	GCGTTCGAAG	TGTCGACGAA
2601	AATCGACAGC	CGCACGATTC	TCGACCAAAT	GGGCGCGGAA	AACCTGCTCG
2651	GTCAGGGCGA	TATGCTGTTT	CTGCCGCCGG	GTAATGCCTA	TCCGCAGCGC
2701	GTTACGCGCG	CGTTTGCCCT	GGATGAAGAG	GTGCACCGCG	TGGTCGAATA
2751	TCTGAAGCAG	TTTGCGGAGC	CGGACTATGT	TGACGATATT	TTGAGCGGCG
2801	GCGGCAGCGA	AGAGCTGCCC	GGCATCGGGC	GCAGCGGCGA	CGGCGAAACC
2851	GATCCGATGT	ACGACGAGGC	CGTATCCGTT	GTCCTGAAAA	CGCGCAAAGC
2901	CAGCATTTTC	GGCGTACAGC	GCGCCTTGCG	CATCGGCTAC	AACCGCGCCG
2951	CGCGTCTGAT	TGACCAAATG	GAAGCGGAAG	GCATTGTGTC	CGCACCGGAA
3001	CACAACGGCA	ACCGTACGAT	TCTCGTCCCC	TTGGACAATG	CTTGA

This corresponds to the amino acid sequence <SEQ ID 496; ORF58ng-1>:

1	MFIVIVLIVV	LLALAGLFFV	RAQSEREWMR	EVS AWQEKKG	EKQAE LPEIK
51	DGMPDFPEFS	LMFLHAVKTA	VYWL FVG VVR	FCRNYLAHES	EPDRPVPPAS
101	ANRADVPTAS	DGYSDSGNGT	EEAETEAAEA	AEEEEADTED	IATAVIDNRR
151	IPFDRSIAEG	LMQSESKTSP	VRPVFKEITL	EEATRALSSA	ALRETKKRYI
201	DAFEKNGTAV	PKVRVSDTPM	EGLQIIGLDD	PVLQRTYSRM	FDADKEAFSE
251	SADYGFEPIYF	EKQHPSAFSA	VKAENARNAP	FRRHAGQEK	QAEAKSPDVS
301	QGQSVSDGTA	VRDARRRVSV	NLKEPNKATV	SAEARISRLI	PESRTVVGKR
351	DVEMPSETEN	VFTETVSSVG	YGGPVYDEAA	DIHIEEPAAP	DAWVVEPPEV
401	PEVAVPEIDI	LPPPPVSEIY	NRTYEPPAGF	EQAQRSRIAE	TDHLAADVLN
451	GGWQEETA AI	ADDGSEGAAE	RSSGOYLSET	EAFGHDSQAV	CPFEDVP SER
501	PSCRVSDETA	DEGAFQSEET	GAVSEHLPTT	DL LPP LFPN	EATQTEEELL
551	ENSITIEEKL	AEFKVKVKVV	DSYSGPVITR	YEIEPDVGVR	GNV LNL EKD
601	LARSLGVASI	RVVETIPGKT	CMGLELPNPK	RQMIRLSEIF	NSPEFAESKS
651	KLTLALGQDI	TGQPVVTDLG	KAPHL LVAGT	TGSGKSVGVN	AMILSMLFKA
701	APEDVRMIMI	DPKMLELSIY	EGITHLLAPV	VTD MKLAANA	LNWCVNEMEK
751	RYRLMSFMGV	RNLAGFNQKI	AEAAARGEKI	GNPFS LTPDD	PEPLEKL PFI
801	VVVVDEFADL	MMTAGKKIEE	LIARLAQKAR	AAGIHLILAT	QRPSVDVITG
851	LIKANIPTRI	AFQVSSKIDS	RTILDQMGAE	NLLGQGDMLF	LPPGTAYPQR
901	VHGAFA SDEE	VHRVVEYLKQ	FGE PDYVDDI	LSGGGSEELP	GIGRSGDG ET
951	DPMYDEAVSV	VLKTRKASIS	GVQRALRIGY	NRAARLIDQM	EAEGIVSAPE
1001	HNGNRTILVP	LDNA*			

ORF58ng-1 and ORF58-1 show 97.2% identity in 1014 aa overlap:

		10	20	30	40	50	60
	orf58-1.pep	MFWIVLIVILLALAGLFFVRAQSEREW	MREVS	SAWQ	EKKGEKQ	AEIPEIKD	GMPDFPELA
	orf58ng-1	MFWIVLIVIVLLALAGLFFVRAQSEREW	MREVS	SAWQ	EKKGEKQ	AEIPEIKD	GMPDFPEFS
5		10	20	30	40	50	60
	orf58-1.pep	LMLFHAVKTAVYWL	FVGVVRF	CRNYLA	HESEPDR	VPVPPAS	ANRADVPTASDGYSDSGNGT
	orf58ng-1	LMLFHAVKTAVYWL	FVGVVRF	CRNYLA	HESEPDR	VPVPPAS	ANRADVPTASDGYSDSGNGT
10		70	80	90	100	110	120
	orf58-1.pep	LMLFHAVKTAVYWL	FVGVVRF	CRNYLA	HESEPDR	VPVPPAS	ANRADVPTASDGYSDSGNGT
	orf58ng-1	LMLFHAVKTAVYWL	FVGVVRF	CRNYLA	HESEPDR	VPVPPAS	ANRADVPTASDGYSDSGNGT
		70	80	90	100	110	120
	orf58-1.pep	EEAETEEAAEEEAAD	TEDIATA	VIDNRR	IPFDRS	IAEGLMP	SESEISPV
	orf58ng-1	EEAETEEAAEEEAAD	TEDIATA	VIDNRR	IPFDRS	IAEGLMQ	SESEISPV
15		130	140	150	160	170	180
	orf58-1.pep	EEAETEEAAEEEAAD	TEDIATA	VIDNRR	IPFDRS	IAEGLMP	SESEISPV
	orf58ng-1	EEAETEEAAEEEAAD	TEDIATA	VIDNRR	IPFDRS	IAEGLMQ	SESEISPV
		130	140	150	160	170	180
	orf58-1.pep	EEATRALNSAALRE	TKKRYI	DAFEK	NETAV	PKVRVS	DTMPEGLQ
	orf58ng-1	EEATRALSSAALRE	TKKRYI	DAFEK	NETAV	PKVRVS	DTMPEGLQ
20		190	200	210	220	230	240
	orf58-1.pep	EEATRALNSAALRE	TKKRYI	DAFEK	NETAV	PKVRVS	DTMPEGLQ
	orf58ng-1	EEATRALSSAALRE	TKKRYI	DAFEK	NETAV	PKVRVS	DTMPEGLQ
		190	200	210	220	230	240
	orf58-1.pep	FDADKEAFSESAD	YGFE	PYFEK	QHPSA	FS	SAVKAENARNAPF
	orf58ng-1	FDADKEAFSESAD	YGFE	PYFEK	QHPSA	FS	SAVKAENARNAPF
25		250	260	270	280	290	300
	orf58-1.pep	FDADKEAFSESAD	YGFE	PYFEK	QHPSA	FS	SAVKAENARNAPF
	orf58ng-1	FDADKEAFSESAD	YGFE	PYFEK	QHPSA	FS	SAVKAENARNAPF
		250	260	270	280	290	300
	orf58-1.pep	QGQSVSDGTAVR	DARRRV	SVNLKE	PNKAT	VSAE	ARISRLIPES
	orf58ng-1	QGQSVSDGTAVR	DARRRV	SVNLKE	PNKAT	VSAE	ARISRLIPES
30		310	320	330	340	350	360
	orf58-1.pep	QGQSVSDGTAVR	DARRRV	SVNLKE	PNKAT	VSAE	ARISRLIPES
	orf58ng-1	QGQSVSDGTAVR	DARRRV	SVNLKE	PNKAT	VSAE	ARISRLIPES
35		310	320	330	340	350	360
	orf58-1.pep	VFTETVSSVGYG	GPVYDE	TA	DIHIEE	PAAPDA	WVVEPPEV
	orf58ng-1	VFTETVSSVGYG	GPVYDE	TA	DIHIEE	PAAPDA	WVVEPPEV
40		370	380	390	400	410	420
	orf58-1.pep	VFTETVSSVGYG	GPVYDE	TA	DIHIEE	PAAPDA	WVVEPPEV
	orf58ng-1	VFTETVSSVGYG	GPVYDE	TA	DIHIEE	PAAPDA	WVVEPPEV
		370	380	390	400	410	420
	orf58-1.pep	NRTYEPPSGFEQ	QRSRIA	ETDHLA	DDVLNG	GWQEET	AAIADDG
	orf58ng-1	NRTYEPPAGFEQ	QRSRIA	ETDHLA	DDVLNG	GWQEET	AAIADDG
45		430	440	450	460	470	480
	orf58-1.pep	NRTYEPPSGFEQ	QRSRIA	ETDHLA	DDVLNG	GWQEET	AAIADDG
	orf58ng-1	NRTYEPPAGFEQ	QRSRIA	ETDHLA	DDVLNG	GWQEET	AAIADDG
		430	440	450	460	470	480
	orf58-1.pep	EAFGHDSQAVCP	FEVNP	SPSR	CRVSD	TEA	DEGAF
	orf58ng-1	EAFGHDSQAVCP	FEVNP	SPSR	CRVSD	TEA	DEGAF
50		490	500	510	520	530	540
	orf58-1.pep	EAFGHDSQAVCP	FEVNP	SPSR	CRVSD	TEA	DEGAF
	orf58ng-1	EAFGHDSQAVCP	FEVNP	SPSR	CRVSD	TEA	DEGAF
		490	500	510	520	530	540
	orf58-1.pep	EATQTEEELENS	ITIEE	KLA	EAFK	VKV	VVDSYSGP
	orf58ng-1	EATQTEEELENS	ITIEE	KLA	EAFK	VKV	VVDSYSGP
55		550	560	570	580	590	600
	orf58-1.pep	EATQTEEELENS	ITIEE	KLA	EAFK	VKV	VVDSYSGP
	orf58ng-1	EATQTEEELENS	ITIEE	KLA	EAFK	VKV	VVDSYSGP
		550	560	570	580	590	600
	orf58-1.pep	LARSLGVASIR	VVETIP	KGKTC	MGLEL	PNPKR	QMIRLSE
	orf58ng-1	LARSLGVASIR	VVETIP	KGKTC	MGLEL	PNPKR	QMIRLSE
60		610	620	630	640	650	660
	orf58-1.pep	LARSLGVASIR	VVETIP	KGKTC	MGLEL	PNPKR	QMIRLSE
	orf58ng-1	LARSLGVASIR	VVETIP	KGKTC	MGLEL	PNPKR	QMIRLSE
65		610	620	630	640	650	660
	orf58-1.pep	TGQPVVTDLGK	APHL	L	VAGTT	SGSGS	VG
	orf58ng-1	TGQPVVTDLGK	APHL	L	VAGTT	SGSGS	VG
70		670	680	690	700	710	720
	orf58-1.pep	TGQPVVTDLGK	APHL	L	VAGTT	SGSGS	VG
	orf58ng-1	TGQPVVTDLGK	APHL	L	VAGTT	SGSGS	VG
		670	680	690	700	710	720

		730	740	750	760	770	780
	orf58-1.pep	EGIPHLLAPVVTDMKLAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
5	orf58ng-1	EGITHLLAPVVTDMKLAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKI					
		730	740	750	760	770	780
	orf58-1.pep	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
10	orf58ng-1	GNPFSLTPDDPEPLEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILAT					
		790	800	810	820	830	840
	orf58-1.pep	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLPPGTAYPQR					
15	orf58ng-1	QRPSVDVITGLIKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLPPGTAYPQR					
		850	860	870	880	890	900
	orf58-1.pep	VHGAFASDEEVHRVVEYLKQFGEPTYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSV					
20	orf58ng-1	VHGAFASDEEVHRVVEYLKQFGEPTYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSV					
		910	920	930	940	950	960
	orf58-1.pep	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVPLDNAX					
25	orf58ng-1	VLKTRKASISGVQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVPLDNAX					
		970	980	990	1000	1010	

30 Furthermore, ORF58ng-1 shows significant homology to the *E.coli* protein FtsK:

	sp P46889 FTSK_ECOLI CELL DIVISION PROTEIN FTSK >gi 1651412 gnl PID d1015290 (D1 division protein FtsK [Escherichia coli] >gi 1651418 gnl PID d1015296 (D90727) Cell division protein FtsK [Escherichia coli] >gi 1787117 (AE000191) cell division protein FtsK [Escherichia coli] Length = 1329
35	Score = 576 bits (1469), Expect = e-163 Identities = 301/459 (65%), Positives = 353/459 (76%), Gaps = 5/459 (1%)
	Query: 556 IEKLAEFKVKVQVDSYSGPVITRYEIEPDVGVGRNSVLNLEKDLARSLGVASIRVVET 615 +E +LA+F++K VV+ GPVITR+E+ GV+ + NL +DLARSL ++RVVE
40	Sbjct: 868 VEARLADFRKADVVNYSPGPVITRFELNLAGPVKAARISNLSRDLARSLSTVAVRVVEV 927
	Query: 616 IPGKTCMGLELNPKNRQIRLSEIFNSPEFAESKSKLTALGQDITGQPVVTDLGKAPHL 675 IPGK +GLELPN KRQ + L E+ ++ +F ++ S LT+ LG+DI G+PVV DL K PHL
45	Sbjct: 928 IPGKPYVGLLEPNKKRQTVYLREVLDNAKFRDNPSPLTVVLGKDIAAGEPVVADLAKMPL 987
	Query: 676 LVAGTTGSGKSVGVNAMILSMLFKAAPEDVRMIMIDPKMLELSIYEGITHLLAPVVTDMK 735 LVAGTTGSGKSVGVNAMILSML+KA PEDVR IMIDPKMLELS+YEGI HLL VVTDMK
50	Sbjct: 988 LVAGTTGSGKSVGVNAMILSMLYKAQPEDVRFIMIDPKMLELSVYEGIPHLLEVTVTDK 1047
	Query: 736 LAANALNWCVNEMEKRYRLMSFMGVRNLAGFNQKIAEAAARGEKIGNPFSLTPDDPEP-- 793 AANAL WCVNEM+RY+LMS +GVRNLAG+N+KIAEA I +P+ D +
55	Sbjct: 1048 DAANALRWCVNEMERRYKLMSALGVRNLAGYNEKIAEADRMMPIDPYWKPGDSMDAQH 1107
	Query: 794 --LEKLPFIVVVVDEFADLMMTAGKKIEELIARLAQKARAAGIHLILATQRPSVDVITGL 851 L+K P+IVV+VDEFADLMMT GKK+EELIARLAQKARAAGIHL+LATQRPSVDVITGL
60	Sbjct: 1108 PVLKKEPYIVVLVDEFADLMMTVGKKVEELIARLAQKARAAGIHLVLTATQRPSVDVITGL 1167
	Query: 852 IKANIPTRIAFQVSSKIDSRTILDQMGAEENLLGQGDMLFLPPGTAYPQRVHGAFASDEEV 911 IKANIPTRIAF VSSKIDSRTILDQ GAE+LLG GDML+ P + P RVHGAF D+EV
65	Sbjct: 1168 IKANIPTRIAFTVSSKIDSRTILDQAGAESLLGMGDMLYSGPNSTLPVRVHGAFVRDQEV 1227
	Query: 912 HRVVEYLKQFGEPTYVDDILSGGGSEELPGIGRSGDGETDPMYDEAVSVVLKTRKASISG 971 H VV+ K G P YVD I S SE G G G E DP++D+AV V + RKASISG
	Sbjct: 1228 HAVVQDWKARGRPQYVDGITSDESEGGAG-GFDGAEELDPLFDQAVQFVTEKRKASISG 1286
	Query: 972 VQRALRIGYNRAARLIDQMEAEGIVSAPEHNGNRTILVP 1010 VQR RIGYNRAAR+I+QMEA+GIVS HNGNR +L P
	Sbjct: 1287 VQRQFRIGYNRAARIIEQMEAQGIVSEQGHNGNREVLAP 1325

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 59

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 497>:-

```

5      1  ATGATTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
      51  CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
     101  TGCTCGGCCG TGCCGCCGAC GGGC..GTGA TCGCCATCGA TGCCGTGTTG
     151  GCATTGGTCG GCTTCTGGGT C.....
//
10     901  .....A TTGCCATCGG TTTGTTTTTA ATTTACCAA ACGGGCTGAC
     951  CCTGCTTTT GAAGCCGTGG AAGACGGCAA AATCCATTT TGGCTCGGAC
    1001  TGCTGCCTAT GCACATTATC ATGTTTGTCC TTGCACTCAT CCTGTGCGC
    1051  GTCCGCAGTA TGCCAGCCA GCCCTCTGG CAGGCGGTTG GCAAAAGTCT
    1101  GACATTGAAA GCGGAAAAAT GA

```

15 This corresponds to the amino acid sequence <SEQ ID 498; ORF101>:

```

      1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GXVIAIDAVL
     51  ALVGFVW... ..
//
    301  ...IAIGLFL IYQNLTLF EAVEDGKIHF WLGLLPMHII MFVLALILLR
    351  VRSMPSQPFW QAVGKSLTLK GK*

```

Further work revealed the complete nucleotide sequence <SEQ ID 499>:

```

      1  ATGATTATC AAAGAAACCT CATCAAAGAA CTCTCTTTTA CCGCCGTCGG
     51  CATTTTCGTC GTCCTCTTGG CGGTATTGGT CTCCACGCAG GCAATCAACC
    101  TGCTCGGCCG TGCCGCCGAC GGGCGTGTG CCATCGATGC CGTGTGGCA
    151  TTGGTCGGCT TCTGGGTCAT CCGTATGACG CCGCTTTTGC TGGTGTGAC
    201  CGCATTATC AGTACGTTGA CCGTGTGAC CCGCTACTGG CGCGACAGCG
    251  AAATGTCGGT CTGGCTATCC TGCGGATTGG CATTGAAACA ATGGATACGC
    301  CCGGTGATGC AGTTTGCCGT GCCGTTGGC GTTTTGGTTG CCGTCATGCA
    351  GCTTTGGGTG ATACCGTGGG CAGAGCTACG CAGCCGCGAA TACGCTGAAA
    401  TCCTGAAGCA GAAGCAGGAA TTGTCTTTGG TGGAGGCAGG CGAGTTCAAC
    451  AGTTTGGGCA AGCGCAACGG CAGGGTTTAT TTTGTGAAA CCTTCGATAC
    501  CGAATCCGGC ATCATGAAAA ACCTGTTCCT GCGCGAACAG GACAAAAACG
    551  GCGGCGACAA CATCATCTTC GCCAAAGAAG GTAACCTCTC GCTGAACGAC
    601  AACAAACGCA CGCTCGAATT GCGCCACGGC TACCGTTACA GCGGCACGCC
    651  CGGACGCGCC GACTACAATC AGGTTTCCTT CCAAAAATC AACCTGATTA
    701  TCAGCACACG GCCCAAATC ATCGACCCCG TTTCCACCG CCGTACCATT
    751  CCGACCCGCC AACTGATTGG CAGCAGCAAC CCGCAACATC AGGCGGAATT
    801  GATGTGGCGC ATCTCGCTGA CCGTCAGCGT CCTCCTACTC TGCCTGCTTG
    851  CCGTGCCGCT TTCCTATTTC AATCCGCGCA GCGGACATAC CTACAATATC
    901  TTGATTGCCA TCGGTTTGT TTTAATTAC CAAAACGGGC TGACCCTGCT
    951  TTTTGAAGCC GTGGAAGACG GCAAATCCA TTTTGGCTC GGACTGCTGC
   1001  CTATGCACAT TATCATGTTT GCCGTTGCAC TCATCCTGTT GCGCGTCCGC
   1051  AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
   1101  GAAAGGCGGA AAATGA

```

45 This corresponds to the amino acid sequence <SEQ ID 500; ORF101-1>:

```

      1  MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
     51  LVGFVIGMT PLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
    101  PVMQFAVPFA VLVAVMQLWV IPWAEILRSRE YAEILKQKE LSLVEAGEFN
    151  SLGKRNGRVY FVETFDTEG IMKNLFLREQ DKNGGDNIIF AKEGNFSLND
    201  NKRTLELRHG YRYSPTGRA DYNQVSFOKL NLIISTTPKL IDPVSHRRTI
    251  PTAQLIGSSN PQHQAEIMWR ISLTVSVLLL CLLAVPLSYF NPSRGHTYNI
    301  LIAIGLFLIY QNLTLLEFA VEDGKIHFVW GLLPMHIIIF AVALILLRVR
    351  SMPSQPFQVA VGKSLTLKGG K*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF101 shows 91.2% identity over a 57aa overlap and 95.7% identity over a 69aa overlap with an ORF (ORF101a) from strain A of *N. meningitidis*:

5	orf101.pep	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFVWX -	10	20	30	40	50
	orf101a	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGXAADXRX-AIDAVLALVGFVXXM	10	20	30	40	50
10	orf101.pepIAIGLFLIYQNGLTLLFEAVEDGKIHFVWLGL	90	100	110		
	orf101a	LTVSVLLLCLLAVPLSYFNPRSGHTYINILXAIGLFLIYQNGLTLLFEAVEDGKIHFVWLGL	280	290	300	310	320
15	orf101.pep	LPMHIIMFVLALILLRVRSMPSQPFVQAVGKSLTLKGGKX	120	130	140	150	
	orf101a	LPMHIIMFVIAIVLLRVRSMPSQPFVQAVGKSLTLKGGKX	340	350	360	370	

The complete length ORF101a nucleotide sequence <SEQ ID 501> is:

1	ATGATTTATC	AAAGAAACCT	CATCAAAGAA	CTCTCTTTTA	CCGCCGTCGG
51	CATTTTCGTC	GTCTCTTGG	CGGTATTGGT	CTCCACGCAG	GCAATCAACC
101	TGCTCGGCCN	TGCCGCCGAC	NGCGTNTCG	CCATCGATGC	CGTGTGGCA
151	TTGGTCGGCT	TCTGGGTCNN	NNGNATGACG	CCGCTTTTGC	TNGTGTGAC
201	CGCATTATC	AGTACGTTGA	CCGTGTTGAC	CCGCTACTGG	CGNACAGCG
251	AAATGTCGGT	CTGGNTATCC	TGCCGATTGG	CATTGAAACA	ATGGATACGC
301	CCGGTGATGC	AGTTTGCCGT	GCCGTTTGCC	GTTTGTTGTTG	CCGTATGCA
351	GCTTTGGGTG	ATACCGTGGG	CAGAGCTACG	CAGCCGCGAA	TACGCTGAAA
401	TCCTGAAGCA	GAAGCAGGAA	TTGTCTTTGG	TGGAGGCAGG	CGGGTCAAC
451	AGTTTGGGCA	AGCGCAACGG	CAGGGTTTAT	TTTGTCGAAA	CCTTCGATAC
501	CGAATCCGGC	ATCATGAAAA	ACCTGTTCCCT	GCGCGAACAG	GACAAAAACG
551	GCGGCGACAA	CATCATCTTC	NCCAAAGAAA	GTAACCTTCTC	GCTGAACGAC
601	AACAAACGCA	CGCTCGAATT	GCGCCACGGC	TACCGTTACA	GCGGCACGCC
651	CGGACGCGCC	GACTACAATC	AGGTTTCCTT	CCNAAAACCTC	AACCTGATTA
701	TCAGCACCA	GCCCCAAACTC	ATCGACCCCG	TTTCCCACCG	CCGTACNATN
751	CCNACNGCCC	AACTGATTGG	CAGCAGCAAC	CCGCAACATC	ANGCGGAATT
801	GATGTGGCGC	ATCTCGCTGA	CCGTCAGCGT	CCTCCTACTC	TGCCTGCTTG
851	CCGTGCCGCT	TTCTTATTTT	AACCCGCGCA	GCGGACATAC	CTACAATATC
901	TTGANTGCCA	TCGGTTTGT	TTTAATTTAC	CAAAACGGGC	TGACCTGCT
951	TTTTGAAGCC	GTGGAAGACG	GCAAATCCA	TTTTGGCTC	GGACTGCTGC
1001	CTATGCACAT	CATCATGTTT	GTCATCGCAA	TCGTACTTCT	GCGCGTCCGC
1051	AGCATGCCCA	GCCAGCCCTT	CTGGCAGGCG	GTTGGCAAAA	GTCTGACATT
1101	GAAAGGCGGA	AAATGA			

This encodes a protein having amino acid sequence <SEQ ID 502>:

1	MIYQRNLIKE	LSFTAVGIFV	VLLAVLVSTQ	AINLLGXAAD	XRXAIDAVLA
51	LVGFVXXMT	PLLLVLTAFI	STLTVLTRYW	RDSEMSVWXS	CGLALKQWIR
101	PVMQFAVPFA	VLVAVMQLWV	IPWAEIERSRE	YAEILKQKQE	LSLVEAGGFN
151	SLGKRNGRVY	FVETFDTESG	IMKNLFLREQ	DKNGGDNIIF	XKESNFSLND
201	NKRTLELRHG	YRYSGTPGRA	DYNQVSFXKL	NLIISTTPKL	IDPVSHRRTX
251	PTAQLIGSSN	POHXAELMWR	ISLTVSVLLL	CLLAVPLSYF	NPRSGHTYNI
301	LXAIGLFLIY	QNGLTLLFEA	VEDGKIHFVW	GLLPMHIIMF	VIAIVLLRVR
351	SMPSPQPFVQA	VGKSLTLKGG	K*		

ORF101a and ORF101-1 show 95.4% identity in 371 aa overlap:

55	orf101a.pep	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGXAADXRXAIDAVLALVGFVXXMT	60
	orf101-1	MIYQRNLIKELSFTAVGIFVVLAVLVSTQAINLLGRAADGRVAIDAVLALVGFVIGMT	60
60	orf101a.pep	PLLLVLTAFISTLTVLTRYWRDSEMSVWXSGLALKQWIRPVMQFAVPFAVLVAVMQLWV	120
	orf101-1	PLLLVLTAFISTLTVLTRYWRDSEMSVWLSGLALKQWIRPVMQFAVPFAVLVAVMQLWV	120

	orf101a.pep	IPWAE LRSREYAEILKQKQELSLVEAGGFNSLGKRN GRVYFVETFDTESGIMKNLFLREQ	180
	orf101-1	IPWAE LRSREYAEILKQKQELSLVEAGGFNSLGKRN GRVYFVETFDTESGIMKNLFLREQ	180
5	orf101a.pep	DKNGGDNIIFXKESNFS LNDNKRTL ELRHGYRYS GTPGRADYNQVSFXKL NLIISTTPKL	240
	orf101-1	DKNGGDNIIFAKEGNFS LNDNKRTL ELRHGYRYS GTPGRADYNQVSFQKL NLIISTTPKL	240
10	orf101a.pep	IDPVSHRRTXPTAQLIGSSNPQHXAELMWRISLTVSVLLLCLLAVPLSYFNPRSGHTYNI	300
	orf101-1	IDPVSHRRTIPTAQLIGSSNPQHQAELMWRISLTVSVLLLCLLAVPLSYFNPRSGHTYNI	300
15	orf101a.pep	LXAIGLFLIYQNGLTLLFEAVEDGKIHFWLGLLPMHIIMFVIAIVLLRVRSMPSQPFWQA	360
	orf101-1	LIAIGLFLIYQNGLTLLFEAVEDGKIHFWLGLLPMHIIMFAVALILLRVRSMPSQPFWQA	360
20	orf101a.pep	VGKSLTLKGGK	371
	orf101-1	VGKSLTLKGGK	371

Homology with a predicted ORF from *N.gonorrhoeae*

ORF101 shows 96.5 % identity in 57aa overlap at the N-terminal domain and 95.1% identity in 61aa overlap at the C-terminal domain, respectively, with a predicted ORF (ORF101ng) from *N.*

25 *gonorrhoeae*:

	orf101.pep	MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGXVIAIDAVLALVGFWV	57
	orf101ng	MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRV-AIDAVLALVGFWVIGM	59
30		//	
	orf101.pep	IAIGLFLIYQNGLTLLFEAVEDGKIHFWLG	333
	orf101ng	SLTVSVLLLCLLAVPLSYFNPRSGHTYNILIAIGLFLIYQNGLTLLFEAVEDGKIHFWLG	331
35	orf101.pep	LLPMHIIMFVLALILLRVRSMPSQPFWQAVGKSLTLKGGK	373
	orf101ng	LLPMHIIMFVIAIVLLRVRSMPSQPFWQAVG	362

The ORF101ng nucleotide sequence <SEQ ID 503> is predicted to encode a protein having partial

40 amino acid sequence <SEQ ID 504>:

	1	MIYQRNLIKE	LSFTAVGIFV	VLLAVLVSTQ	AINLLGRAAD	GRVAIDAVLA
	51	LVGEFWIGMT	PLLLVLTAFI	STLTVLTRYW	RDSEMSVWLS	CGLALKQWIR
	101	PVMQFAVPFA	ILIAVMQLWV	IPWAE LRSRE	YAEILKQKQE	LSLVEAGEFN
	151	NLGKRN GRVY	FVETFDTESG	IMKNLFLREQ	DKNGGDNIIF	AKEGNFSLKD
45	201	NKRTL ELRHG	YRYS GTPGRA	DYNQVSFQKL	NLIISTTPKL	IDPVSHRRTI
	251	STAQLIGSSN	PQHQAELMWR	ISLTVSVLLL	CLLAVPLSYF	NPRSGHTYNI
	301	LIAIGLFLIY	QNGLTLLFEA	VEDGKIHFWL	GLLPMHIIMF	VIAIVLLRVR
	351	SMPSQPFWQA	VG...			

Further work revealed the complete nucleotide sequence <SEQ ID 505>:

50	1	ATGATTATC	AAAGAAACCT	CATCAAAGAA	CTCTCTTTTA	CGCCCGTCGG
	51	CATTTTCGTC	GTCCTCTTGG	CGGTGTTGGT	GTCCACGCAG	GCGATCAACC
	101	TGCTTGCCG	CGCAGCTGAC	GGGCGTGTG	CCATCGATGC	CGTGTGGCC
	151	TTAGTCGGCT	TCTGGGTCAT	CGGTATGACC	CCGCTTTTGC	TGGTGTGAC
	201	CGCATTATC	AGCACGCTGA	CGTATATGAC	CCGCTACTGG	CGCGACAGCG
55	251	AAATGTCGGT	CTGGCTATCC	TGCGGATTGG	CGTTGAAACA	GTGGATACGC
	301	CCCGTCATGC	AGTTTGCCGT	GCCGTTTGCC	ATCCTGATTG	CCGTATGCA
	351	GCTTTGGGTG	ATACCGTGGG	CAGAGCTGCG	CAGCCGCGAA	TATGCCGAAA
	401	TTTGAAGCA	GAAGCAGGAA	TTGTCTTTGG	TGGAAGCCGG	CGAGTTCAAT
	451	AAC TTGGGCA	AGCGCAACGG	CAgggtttaT	Ttcgtcgaaa	CCTTTGACAC
60	501	CGaatccgGC	ATCATGAAAA	ACCTGTtctt	GcGCGAACAG	GACAAAAACG
	551	gcggcgacaA	CATCATCTTC	GCaaaGAag	gtaactTctc	gctgaaggac

-298-

5
601 AACAAAcgca cgctcgaATT GCGCCACGGC TACCGTTACA GCGGcacgcC
651 CGGacGCGCc gactaCAATC AGGTTtcctt cCAAAAacTc aacctgATta
701 TCAGCACCAC GCCCAAacTT ATCGaccCCG TTTCCACCCG CCGCACCATT
801 GATGTGGCGC ATCTCGTGGA CAGCAGCAAT CCGCAACATC AGGCAGAATT
851 CCGTGCCGCT TTCCTATTTT AACCCGCGCA GCGGACATAC CTACAATATC
901 TTGATTGCCA TCGGTTTGT TTAATTTAC CAAAACGGGC TGACCCTGCT
951 TTTTGAAGCC GTGGAAGACG GCAAAATCCA TTTTGGCTC GGACTGCTGC
10
1001 CTATGCACAT CATCATGTTC GTCATCGCAA TCGTACTTCT GCGCGTCCGC
1051 AGTATGCCCA GCCAGCCCTT CTGGCAGGCG GTTGGCAAAA GTCTGACATT
1101 GAAAGgcgGA AAATGA

This corresponds to the amino acid sequence <SEQ ID 506; ORF101ng-1>:

15
1 MIYQRNLIKE LSFTAVGIFV VLLAVLVSTQ AINLLGRAAD GRVAIDAVLA
51 LVGFVWIGMT PLLLVLTAFI STLTVLTRYW RDSEMSVWLS CGLALKQWIR
101 PVMQFAVPFA ILIAVMQLWV IPWAEIRSRE YAEILKQKE LSLVEAGEFN
151 NLGKRNGRVY FVETFDTESG IMKNLFLREQ DKNNGDNIIIF AKEGNFSLKD
201 NKRTLELRHG YRYSCTPGRA DYNQVSFQKL NLIISTPKL IDPVSHRRTI
251 STAQLIGSSN PQHQAEMLWR ISLTVSVLLL CLLAVPLSYF NPRSGHTYNI
301 LIAIGLFLIY QNGLTLLFEA VEDGKIHFVW GLLPMHIIMF VIAIVLLRVR
20
351 SMPSPFWQA VGKSLTLKGG K*

ORF101ng-1 and ORF101-1 show 97.6% identity in 371 aa overlap:

		10	20	30	40	50	60
25	orf101-1.pep	MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFVWIGMT					
	orf101ng-1	MIYQRNLIKELSFTAVGIFVLLAVLVSTQAINLLGRAADGRVAIDAVLALVGFVWIGMT					
		10	20	30	40	50	60
30	orf101-1.pep	PLLLVLTAFISTLTVLTRYWRDSEMSVWLS CGLALKQWIRPVMQFAVPFAVLVAVMQLWV					
	orf101ng-1	PLLLVLTAFISTLTVLTRYWRDSEMSVWLS CGLALKQWIRPVMQFAVPFAVLVAVMQLWV					
		70	80	90	100	110	120
35	orf101-1.pep	IPWAEIRSREYAEILKQKQELSLVEAGEFNSLGRNGRVYFVETFDTESGIMKNLFLREQ					
	orf101ng-1	IPWAEIRSREYAEILKQKQELSLVEAGEFNNLGRNGRVYFVETFDTESGIMKNLFLREQ					
		130	140	150	160	170	180
40	orf101-1.pep	DKNNGDNIIIFAKEGNFSLNDNKRTLELRHG YRYSCTPGRADYNQVSFQKL NLIISTPKL					
	orf101ng-1	DKNNGDNIIIFAKEGNFSLKDNKRTLELRHG YRYSCTPGRADYNQVSFQKL NLIISTPKL					
		190	200	210	220	230	240
45	orf101-1.pep	IDPVSHRRTIPTAQLIGSSNPQHQAELMWRISLTVSVLLLC LLAVPLSYFNPRSGHTYNI					
	orf101ng-1	IDPVSHRRTISTAQLIGSSNPQHQAELMWRISLTVSVLLLC LLAVPLSYFNPRSGHTYNI					
		250	260	270	280	290	300
50	orf101-1.pep	LIAIGLFLIYQNGLTLLFEAVEDGKIHFVW GLLPMHIIMFAVALILLRVRSMPSQPFWQA					
	orf101ng-1	LIAIGLFLIYQNGLTLLFEAVEDGKIHFVW GLLPMHIIMFVIAIVLLRVRSMPSQPFWQA					
		310	320	330	340	350	360
55	orf101-1.pep	VGKSLTLKGGKX					
	orf101ng-1	VGKSLTLKGGKX					
		370					
60							

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is

predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 60

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 507>:

```

5      1  ..GGTGGTGGTT TTATCAATGC TTCCTGTGCC ACTTTGACGA CAGCCAAACC
      51  GCAATATCAA GCAGGAGACC TTAGCGCTTT TAAGATAAGG CAAGGCAATG
     101  TTGTAATCGC CGGACACGGT TTGGATGCAC GTGATACCGA TTACACACGT
     151  ATTCTCAGTT ATCATTCCAA AATCGATGCA CCCGTATGGG GACAAGATGT
     201  TCGTGTGCTC GCGGGACAAA ACGATGTGGC CGCAACAGGT GATGCACATT
     251  CGCCTATTCT CAATAATGCT GCTGCCAATA CGTCAAACAA TACAGCCAAC
10    301  AACGGCACAC ATATCCCTTT ATTTGCGATT GATACAGGCA AATTAGGAGG
     351  TAT.GTATGC CAACAAAATC ACCTTGATCA GTACGGTCGA GCAAGCAGGC
     401  ATTCGTAA

```

This corresponds to the amino acid sequence <SEQ ID 508; ORF113>:

```

15      1  ..GGGFINASCA TLTTAKPOYQ AGDLSAFKIR QGNVVIAGHG LDARDTDYTR
      51  ILSYHSKIDA PVWGDVVRV AGQNDVAATG DAHSPILNNA AANTSNNNTAN
     101  NGTHIPLFAI DTGKLGXVC QQNHLQYGR ASRHS*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with with pspA putative secreted protein of *N.meningitidis* (accession AF030941)

20 ORF and pspA show 44% aa identity in 179aa overlap:

```

      orf113  GGGFINASCATLTTAKPOYQAGDLSAFKIRQGNVVIAGHGLDARDTDYTRILSYHSKIDA 60
      pspA    GGG INA+ TLT+ P   G+L+ F + G VVI G GLD D DYTRILS ++I+A
      pspA    GGGLINAASVTLTSGVPVLNNGNLTGFDVSSGKVVI GGKGLDTSADYTRILSRAAEINA 256
25      orf113  PVWGDVVRVAGQNDVAATGDAHSPILXXXXXXXXXXXXXGTHIPLFAIDTGKLGMYA 120
      pspA    VWG+DV+VV+G+N + G               + P AIDT LGMYA
      pspA    GVWGDVVKVSGKNKLDG-----SLAKTASAPSSSDSVTPTVAIDTATLGMYA 307
30      orf113  NKITLISTVEQAGIRNQGWFAAGNVAVNAEGKLVNTGMIAATGENHAVSLHARNVHN 179
      pspA    +KITLIST A IRN+G+ FA+ G V ++A+GKL N+G I A      +++ A+ V N
      pspA    DKITLISTDNGAVIRNKGRIFAATGGVTLSDAGKLSNSGSIDAA----EITISAQTVDN 362

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF113 shows 86.5% identity in 52aa overlap at the N- terminal part and 94.1% identity in 17aa

35 overlap at the C-terminal part with a predicted ORF (ORF113ng) from *N. gonorrhoeae*:

```

      orf113                                GGGFINASCATLTTAKPOYQAGDLSAFKIR 30
      orf113ng  SHPSQLNGYIEVGRRRAEVVIANPAGIAVNGGGFINASRATLTGQPOYQAGDFSGFKIR 224
40      orf113  QGNVVIAGHGLDARDTDYTRILSYHSKIDAPVWGDVVRVAGQNDVAATGDAHSPILNNA 90
      orf113ng  QGNAVIAGHGLDARDTDFTRILVCQQNHLQYGRTSRHS 263
      orf113                                IDTGKLGXVCQQNHLQYGRASRHS 135
45      orf113ng  DFGFKIRQGNVVIAGHGLDARDTDFTRILVCQQNHLQYGRTSRHS 263

```

The complete length ORF113ng nucleotide sequence <SEQ ID 509> is predicted to encode a protein having amino acid sequence <SEQ ID 510>:

```

50      1  MNKTLYRVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIFTH
      51  SKAFCFSALG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTGNGIP

```

```

101 QVNIQTPTSA GVSUNQYQF DVGNRGAILN NSRSNTQTQL GGWIQGNPWL
151 TRGEARVVVN QINSSHPSQL NGYIEVGRR AEVVIANPAG IAVNGGGFIN
201 ASRATLTGQ PQYQAGDFSG FKIRQGNVAV AGHGLDARDT DFTRILVCQO
251 NHLDQYGRTS RHS*

```

- 5 Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 61

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 511>:

```

10      1 ..TCAACGGGAC ATAGCGAACA AAATTACACT TTGCCGCGAG AAATCACACG
      51 CAACATTTC A CTGGGTTTCAT TTGCCTATGA ATCGCATCGC AAAGCATTAA
101     101 GCCATCATGC GCCCAGCCAA GGCACGTAGT TGCCGCAAAG CAACGGTATT
151     151 TCGCTACCCT ATACGTCCAA TTCTTTTACC CCATTACCCA GCAGCAGCTT
201     201 ATACATTATC AATCCTGTCA ATAAAGGCTA TCTTGTTGAA ACCGATCCAC
251     251 GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GctGGACAGC
15      301 CTCAAAC TAG ACCCAAACAA TTACATAAA CGTTTGGGTG ATGGTTATTA
351     351 CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
401     401 GTTTAGAcGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT
451     451 AATGGCGCGA CTGCGGCACG TTcGATGAAT CTCAGCGTTG GCATTGCATT
501     501 AAGTGCCGAG CAAGTAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
20      551 AAAAAGAAGT TAAGCTTCCT GATGGCGGCA CACAAACCGT ATTGGTGCCA
601     601 CAGGTTTATG TACGCGTTAA AAATGGCGAC ATAGACGGTA AAGGTGCATT
651     651 GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
701     701 CAGGCACGAT TGCAGGgCGC AATGCGCTTA TTATCAATAC CGATACGCTA
25      751 GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
801     801 ACAAGACATC AATAATATTG GCGGCATGCT TTCTGCCGAA CAGACATTAT
851     851 TGCTCAACGC AGGCAACAAC ATCAACAGCC AAAGCACCAC CGCCAGCAGT
901     901 CAAAATACAC AAGGCAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
951     951 TATCACAGGC AAAGAAAAG GTGTTT..

```

This corresponds to the amino acid sequence <SEQ ID 512; ORF115>:

```

30      1 ..STGHSEQNYT LPREITRNI LGSFAYESH KALSHHAPSQ GTELPQSNGI
      51 SLPYTSNSFT PLPSSSLYII NPVNKGYLE TDPRFANYRQ WLGSDYMLDS
101     101 LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
151     151 NGATAARSMN LSVGIALSAE QVAQLTSDIV WLWQKEVKLP DGGTQTVLVP
201     201 QVYVRVKNGD IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
35      251 DNIGGRIHAQ KSAVTATQDI NNIGGMLSAE QTLNAGNN INSQSTTASS
301     301 QNTQGSSTYL DRMAGIYITG KERGV..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the *pspA* putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF115 and *pspA* protein show 50% aa identity in 325aa overlap:

```

40      Orf115: 1 STGHSEQNYTLPREITRNI LGSFAYESH KALSHHAPSQGTLPQSNGISLPYTSNSFT 60
      STG+S Y E++ +I +G AY+ + + P + NGI +T
      pspA: 778 STGYSRSPYEPAPEVS-SIRMGISAYKGYAPQASDIPGTVPVVAENGIHPTFT----- 831
45      Orf115: 61 PLPSSSLYIINPVNKGYLETDP RFANYRQWL GSDYMLDSL KLDPNNLHKRLGDGYEQR 120
      LP+SSL+ I P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQR+
      pspA: 832 -LPNSSLFAIAPNNKGYLIETDPAFTDYRKWL GSGYMLAALQQDPNHIHKRLGDGYEQRK 890
50      Orf115: 121 LINEQIAELTGHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIV 180
      L+NEQIA+LTG+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQVA+LTSDIV
      pspA: 891 LVNEQIAKLTGYRRLDGYTNDDEEQFKALMDNGITIAKELQTPGIALSAEQVARLTSDIV 950
55      Orf115: 181 WLWQKEVKLPDGGTQTVLVPQVYVRVKN GIDGKGALLSGSNTQINVSGSLKN-SGTIAG 239
      WL + V LPDG TQTVL P+VYVR + D++G+GALLSGS I SG+++N G IAG
      pspA: 951 WLENETVTLPDGTQTQTVLKP VYVRARPKDMNGQALLSGSVVDIG-SGA IENRGGLIAG 1009

```

-301-

Orf115: 240 RNALIINTDTLDNIGGRIHAQKSAVTATQDINNIGGMLSAEQTLNAGXXXXXXXXXX 299
 R ALI+N + N+ G + + A DI N G + AE LLL A
 pspA: 1010 REALILNAQNIKNLQGDQGNIFAAAGSDITNTGS-IGAENALLKASNNIESRSETRS 1068

5 Orf115: 300 XXXXXXXXXXXYLDRMAGIYITGKEG 324
 + R+AGIY+TG++ G
 pspA: 1069 NQNEQGSVRNIGRVAGIYLTGRQNG 1093

Homology with a predicted ORF from *N.gonorrhoeae*

10 ORF115 shows 91.9% identity over a 334aa overlap with a predicted ORF (ORF115ng) from *N.gonorrhoeae*:

	orf115.pep	STGHSEQNYTLPREITRNISLGSFAYESHK	31
	orf115ng	NEQTFGEKKVSENGKLHNYWRARRKGHDETGHRQNYTLPEEITRDISLGSFAYESHK	71
15	orf115.pep	ALSHHAPSQGTLPQSN-----GISLPYTSNSFTPLPSSSLYIINPVNKGVLVET	81
	orf115ng	ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPSSSLYIINPVNKGVLVET	131
20	orf115.pep	DPRFANYRQWLGS DYMLDSLKLDPNNLHKRLGDGYEQRLINEQIAELTGHRRLDGYQND	141
	orf115ng	DPRFANYRQWLGS DYMLGSLKLDPNNLHKRLGDGYEQRLINEQIAELTGHRRLDGYQND	191
25	orf115.pep	EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWLVOKEVKLPDGGTQTVLVPQ	201
	orf115ng	EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLVOKEVKLPDGGTQTVLMPQ	251
	orf115.pep	VYVRVKNGDIDGKGALLSGSNTQINVSGLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	261
30	orf115ng	VYVRVKNGGIDGKGALLSGSNTQINVSGLKNSGTIAGRNALIINTDTLDNIGGRIHAQK	311
	orf115.pep	SAVTATQDINNIGGMLSAEQTLNAGNNINSQSTTASSQNTQGSSTYLDRMAGIYITGK	321
	orf115ng	SAVTATQDINNIGGILSAEQTLNAGNNINNQSTAKSSQNAQGSSTYLDRMAGIYITGK	371
35	orf115.pep	EKGV	325
	orf115ng	EKGVLAAGKIDINIIAGQISNQSDQGQTRLQAGRDINLDTVQTGKYQEIHFADNHTIR	431

An ORF115ng nucleotide sequence <SEQ ID 513> was predicted to encode a protein having amino acid sequence <SEQ ID 514>:

	1	MLVQTEKDGL	HNEQTFGEKK	VFSENGKLHN	YWRARRKGHD	ETGHRQNYT
	51	LPEEITRDIS	LGSFAYESHK	KALSRHAPSQ	GTELPQSNRD	NIRTAKSNGI
	101	SLPYTPNSFT	PLPGSSLYII	NPANKGYLVE	TDPRFANYRQ	WLGS DYMLGS
45	151	LKLDPNNLHK	RLGDGYEQR	LINEQIAELT	GHRRLDGYQN	DEEQFKALMD
	201	NGATAARSMN	LSVGIALSAE	QAAQLTSDIV	WLVOKEVKLP	DGGTQTVLMP
	251	QVYVRVKNGG	IDGKGALLSG	SNTQINVSGL	LKNSGTIAGR	NALIINTDTL
	301	DNIGGRIHAQ	KSAVTATQDI	NNIGGILSAE	QTLNAGNN	INNQSTAKSS
	351	QNAQGSSTYL	DRMAGIYITG	KEKGVLAQA	GKDINIIAGQ	ISNQSDQGQT
50	401	RLQAGRDINL	DTVQTGKYQE	IHFADNHTI	RGSTNEVGSS	IQTKGDVTL
	451	SGNNLNAKAA	EVGSAKGTLA	VYAKNDITIS	SGIHAGQVDD	ASKHTGRSGG
	501	GNKLVIDDKA	QSHHETAQSS	TFEGKQVVLQ	AGNDANILGS	NVISDNTRI
	551	QAGNHVRIGT	TQTQSQSEY	HQTKSGLMS	AGIGFTIGSK	TNTQENQSQS
	601	NEHTGSTVGS	LKGDITIVAS	KHYEQTGSNV	SSPEGNNLIS	TQSM DIGAAQ
	651	NQLNSKTTQT	YEQKGLTVAF	SSPVDLAQQ	AIAVAHAQAK	QFDKAKTTAL
55	701	MPWRLPMQVG	RLFKQAKAPK	K*		

Further work revealed the following partial gonococcal DNA sequence <SEQ ID 515>:

	1	TTGCTTGTGC	AAACAGAAAA	AGACGGTTTG	CATAACGAGC	AAACCTTTGG
	51	CGAGAAGAAA	GTCTTCAGCG	AAAATGGTAA	GTTGCACAC	TACTGGCGTG
	101	CGCGTCGTAA	AGGACATGAT	GAAACAGGGC	ATCGTGAACA	AAATTATACT
60	151	TTGCCGGAGG	AAATCACACG	CGACATTTCA	CTGGGTTTCAT	TTGCCTATGA
	201	ATCGCATAGC	AAAGCATTAA	GCCGTCATGC	GCCCAGCCAA	GGCACTGAGT
	251	TGCCACAAAG	TAACCGGGAT	AATATCCGTA	CTGCGAAAAG	CAACGGTATT

301 TCGCTACCCT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT
 351 ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTGTGA ACCGATCCAC
 401 GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGGCAGC
 451 CTCAAACCTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
 501 CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
 551 GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT
 601 AATGGCGCGA CTGGCGGCAG TTCGATGAAT CTCAGCGTTG GCATTGCATT
 651 AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATTGTT TGGTTGGTAC
 701 AAAAAAGAGT TAAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA
 751 CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT
 801 GTTGTCAGGC AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAAACT
 851 CAGGCACGAT TGCAGGCGCG AATGCGCTTA TTATCAATAC CGATACGCTA
 901 GACAATATCG GTGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
 951 ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT
 1001 TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT
 1051 CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA
 1101 TATCACAGGC AAAGAAAAAG GTGTTTTAGC AGCGCAGGCA GGCAAAGACA
 1151 TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGCAAAACC
 1201 CGGCTGCAGG CAGGACGCGA CATTAACCTG GATACGGTAC AAACCGGCAA
 1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA
 1301 CGAACGAAGT CGGCAGCAGC ATTCAAACAA AAGGCGATGT TACCtATG
 1351 TCAGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAAGG
 1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC
 1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC
 1501 GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAACTGC
 1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
 1601 ATGCCAACAT CCTTGGCAGT AATGTTATTT CCGATAATGG CACCCGGATT
 1651 CAAGCAGGCA ATCATGTTTCG CATTGGTACA ACCCAAATC AAAGCCAAAG
 1701 CGAAACCTAT CATCAAAACC AAAAATCAGG ATTGATGAGT GCAGGTATCG
 1751 GCTTCACTAT TGGCAGCAAG ACAAACACAC AAGAAAACCA ATCCCAAGC
 1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT
 1851 TGTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCAGCCCTG
 1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA
 1951 AACCAATTAA ACAGCAAAAC CACCCAAACC TACGAACAAA AAGGCTTAAC
 2001 GGTGGCATTG AGTTCGCCCG TTACCGATTG GGCACAACAA GCGATTGCCG
 2051 TAGCACACAA AGCAGCAAAC AAGTCGGACA AAGCAAAAC GACCGCGTTA
 2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA
 2151 GGCGCACAAA ACTTAG

This corresponds to the amino acid sequence <SEQ ID 516; ORF115ng-1>:

40 1 LLVQTEKDGL HNEQTFGEKK VSENGKLHN YWRARRKGHD ETGHREQNYT
 51 LPPEITRDIS LGSFAYESH KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
 101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPREFANYRQ WLGS DYMLGS
 151 LKLDPNNLHK RLGDGYYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
 201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVQKEVKLP DGGTQTVLMP
 45 251 QVYVRVKNKG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
 301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLLLNAGNN INNQSTAKSS
 351 QNAQGSSTYL DRMAGIYITG KEGVLAQA GKDINI IAGQ ISNQSDQGQT
 401 RLQAGR DINL DTVOTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTL
 451 SGNNLNAAKAA EVGSAGKTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
 50 501 GNKLVIDKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNTRI
 551 QAGNHVRIGT TQTQSQSEY HQTKSGLMS AGIGFTIGSK TNTQENQSQS
 601 NEHTGSTVGS LKGDITIVAS KHYEQTGSNV SSPEGNNLIS TQSM DIGAAQ
 651 NQLNSKTTQT YEQKGLTVAF SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL
 701 MFWR LPMQVG RPIKQAKAHK T*

55 This gonococcal protein (ORF115ng-1) shows 91.9% identity with ORF115 over 334aa:

20 30 40 50 60 70
 orf115ng-1.p NEQTFGEKKVSENGKLHNYWRARRKGHD ETGHREQNYTLPEEITRDISLGSFAYESHK
 orf115 STGHSEQNYTLPREITRNISLGSFAYESHK
 10 20 30
 80 90 100 110 120 130
 orf115ng-1.p ALSRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYIINPANKGYLVET
 orf115 ALSHHAPSQGTLPQSN-----GISLPYTSNSFTPLPSSSLYIINPVNKG YLVET
 40 50 60 70 80

-303-

```

      140      150      160      170      180      190
orfl15ng-1.p DPRFANYRQWLGS DYMLGSLKLD PNNLHKRLG DGYEQRLINEQIAELTG HRRLDGYQND
5 orfl15      DPRFANYRQWLGS DYMLGSLKLD PNNLHKRLG DGYEQRLINEQIAELTG HRRLDGYQND
      90      100      110      120      130      140

      200      210      220      230      240      250
orfl15ng-1.p EEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWL VQKEVKLPDGGTQT VLMPO
10 orfl15      EEQFKALMDNGATAARSMNLSVGIALSAEQVAQLTSDIVWL VQKEVKLPDGGTQT VLMPO
      150      160      170      180      190      200

      260      270      280      290      300      310
orfl15ng-1.p VYVRVKNGGIDGKGALLSGSNTQINVSGSLKNSGTIAGR NALIINTDTL DNIGGRIHAQK
15 orfl15      VYVRVKNGGIDGKGALLSGSNTQINVSGSLKNSGTIAGR NALIINTDTL DNIGGRIHAQK
      210      220      230      240      250      260

      320      330      340      350      360      370
orfl15ng-1.p SAVTATQDINNIGGILSAEQ TLLNAGNNINNQSTAKSSQNAQGSSTY LDRMAGIYITGK
20 orfl15      SAVTATQDINNIGGILSAEQ TLLNAGNNINNQSTAKSSQNAQGSSTY LDRMAGIYITGK
      270      280      290      300      310      320

      380      390      400      410      420      430
orfl15ng-1.p EKGVLAAQAGKDINIIAGQISNQSDQGQTRLQAGR DINLDTVQTGKYQEIHFDADNHTIR
25 orfl15      EKGVLAAQAGKDINIIAGQISNQSDQGQTRLQAGR DINLDTVQTGKYQEIHFDADNHTIR
      IIII
      EKGVL

```

In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

```

30 gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis] Length
   = 2273
   Score = 604 bits (1541), Expect = e-172
   Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

35 Query: 1 LLVQTEKDGLHNEQTFGEKKVSENGKLHNYWRARRKGHD E TGHREQNYTLPEEITRDIS 60
   L+V T + L N++T G K + ++ G LH Y R +KG D TG+ Y E++ I
Sbjct: 739 LIVGTPESALDND ETLGTKTI-TDKGDLHRYHRHKKGRDSTGYSRSPYEPAPEVS-SIR 796

40 Query: 61 LGSFAYESHKALSRRHAPSQGTLPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120
   +G AY+ + AP Q +++P + + NGI +T LP SSL+ I
Sbjct: 797 MGISAYKGY-----APQQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840

45 Query: 121 NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLD PNNLHKRLG DGYEQRLINEQIAELT 180
   P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLG DGYEQL+L+NEQIA+LT
Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGS GYMLAALQQDPNHHKRLG DGYEQKLVNEQIAKLT 900

50 Query: 181 GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWL VQKEVKLP 240
   G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP
Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQ LTPGIALSAEQVARLTS DIVWLENETVTLP 960

55 Query: 241 DGGTQT VLMPOQVYVRVKNGGIDGKGALLSGSNTQINVSGSLKN-SGTIAGR NALIINTDT 299
   DG TQTVL P+VYVR + ++G+GALLSGS I SG+++N G IAGR ALI+N
Sbjct: 961 DGT TQTVLKP KVVYVRARPKDMNGQGALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019

60 Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQ TLLNAGNNINNQSTAKSSQNAQGSSTY 359
   + N+ G + + A DI N G I AE LLL A NNI ++S +S+QN QGS
Sbjct: 1020 IKNLQGD LQGNIFAAAGSDITNTGSI- GAENALLKASN NIESRSETRSNQNEQGSVRN 1078

65 Query: 360 LDRMAGIYITGKEKGVLAQAGKDINIIAGQISNQSDQGQTRLQAGR DINLDTVQTGKYQ 419
   + R+AGIY+TG++ G + AG +I + A +++NQ S+ GQT L AG DI DT + Q
Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSE DGT VLNAGGDIRSDTTGISRNQ 1138

70 Query: 420 EIHFADNHTIRGSTNEVGSSIQTKGDVTL LSGNNLNAKAAEVGSAKGT LAVYAKNDITI 479
   FD+DN+ IR NEVGS+I+T+G+++L + ++ +AAEVGS +G L + A DI +
Sbjct: 1139 NTIFDSDNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQGR LKLAAGRDIKV 1198

Query: 480 SSGIHAGQVDDASKHTGRSGGKNLKVITDKAQSHHETAQSSTFEKGQVVLQAGNDANILG 539
   +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G
Sbjct: 1199 EAGKAHTETEDALKYTRSGGGIKQKMT RHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258

```

-304-

Query: 540 SNVISDNGTRIQAQNHVRIGTTQTQSQSEYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S
 Sbjct: 1259 SNIIADNHTILSAKNNIVLKAETR SRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNRS 1318

5 Query: 599 QSNEHTGSTVGS LKGD TTVASKHYEQ TGS NVSSPEGNNLISTQSM DIGAAQNQLNSKTT 658
 ++ HT S VGS L G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRYSQESK 1378

10 Query: 659 QTYEQKGLTVAFSSPVT D 676
 Q YEQKG+TVA S PV +
 Sbjct: 1379 QVYEQKGVTVVAISVPVVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 62

15 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 517>:

1 ..TCAGGGAATA ACCTCAATGC CAAAGCTGCC GAAGTCAGCA GCGCAAACGG
 51 TACACTCGCT GTGTCTGCCA ATAATGACAT CAACATCAGC GCAGGCATCA
 101 ACACGACCCA TGTGTATGAT GCGTCCAAAC ACACAGGCAG AAGCGGTGGT
 151 GGCAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACCGC
 201 CCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
 251 ATGCCAACAT CCTTGGCAGC AATGTTATTT CCGATAATGG CACCCAGATT
 301 CAAGCAGGCA ATCATGTTTC CATTTGGTACA ACCCAAATC AAAGCCAAAG
 351 CGAAACCTAT CATCAAACCC AGAAATCAGG ATTGATGAGT GCAGGTATCG
 401 GCTTCACTAT TGGCAGCAAG ACAACACAC AAGAAACCA ATCCCAAAGC
 25 451 AACGAACATA CAGGCAGTAC CGTAGGCAGC TTGAAAGGCG ATACCACCAT
 501 TGTTCAGGC AAACACTACG AACAAATCGG CAGTACCGTT TCCAGCCCGG
 551 AAGGCAACAA TACCATCTAT GCCCAAAGCA TAGACATTCA AGCGGCACAC
 601 AACAAATTAA ACAGTAATAC CACCCAAACC TATGAACAAA AAGG.CTAAC
 651 GGTGCATT C AGTTCGCCCG TTACCGATT GGCACAACAA ...

30 This corresponds to the amino acid sequence <SEQ ID 518; ORF117>:

1 ..SGNNLNAAKAA EVSSANGTLA VSANNDINIS AGINTTHVDD ASKHTGRSGG
 51 GNKLIVITDKA QSHHETAQSS TFEKGQVVLQ AGNDANILGS NVISDNGTQI
 101 QAGNHVRIGT TQTQSQSEY HQTQKSGLMS AGIGFTIGSK TNTQENQSQS
 151 NEHTGSTVGS LKGD TTVIVAG KHYEQIGSTV SSPEGNNIY AQSIDIQAHH
 35 201 NKLNSNTTQT YEQKXLTVAF SSPVTDLAQQ ...

Computer analysis of this amino acid sequence gave the following results:

Homology with the pspA putative secreted protein of *N.meningitidis* (accession number AF030941)

ORF117 and pspA protein show 45% aa identity in 224aa overlap:

Orf117: 4 NLNAAAEVSSANGTLAVSANNDINISAGINTTHVDDASKHTGRSGGGNKLIVITDKAQSH 63
 ++ +AAEV S G L ++A DI + AG T +DA K+TGRSGGG K +T ++
 40 pspA: 1173 DIRIRAAEVGSEQRLKLAAGRDIKVEAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQ 1232

Orf117: 64 HETAQSSTFEKGQVVLQAGNDANILGSNVISDNGTQIQAGNHVRIGTTQTQSQSEYHQT 123
 + A S T +GK+++L +G D + GSN+I+DN T + A N++ + +T+S+S ++
 45 pspA: 1233 NGQAVSGTLDGKEIILVSGRDITVTGSNIIADNHTILSAKNNIVLKAETR SRSAEMNKK 1292

Orf117: 124 QKSGLM-SAGIGFTIGSKTNTQENQSQSNEHTGSTVGS LKGD TTVIVAGKHYEQIGSTVSS 182
 +KSGLM S GIGFT GSK +TQ N+S++ HT S VGS L G+T I AGKHY Q GST+SS
 50 pspA: 1293 EKSGLMGSGGIGFTAGSKKDTQTNRSETVSHTESVVGSLNGNTLISAGKHYTQTGSTISS 1352

Orf117: 183 PEGNNIYAQSIDIQAHHNKLNSNTTQTYEQKXLTVAFSSPVT D 226
 P+G+ I + I I AA N+ + + Q YEQK +TVA S PV +
 pspA: 1353 PQGDVGISSGKISIDAAQNRYSQESKQVYEQKGVTVVAISVPVVN 1396

Homology with a predicted ORF from *N.gonorrhoeae*

ORF117 shows 90% identity over a 230aa overlap with a predicted ORF (ORF117ng) from *N.gonorrhoeae*:

5	orf117.pep	SGNNLNAAAEVSSANGTLAVSANNDINIS	30
	orf117ng	IHFDADNHTIRGSTNEVGSSIQTKGDTVLLSGNNLNAAAEVGSAGKTLAVYAKNDITIS	480
10	orf117.pep	AGINTTHVDDASKHTGRSGGKNLVIDTKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	90
	orf117ng	SGIHAGQVDDASKHTGRSGGKNLVIDTKAQSHHETAQSSTFEGKQVVLQAGNDANILGS	540
15	orf117.pep	NVISDNGTQIQAGNHVRIGTTQTQSQSEYHQTQKSLMSAGIGFTIGSKTNTQENQSQS	150
	orf117ng	NVISDNGTRIAGNHVRIGTTQTQSQSEYHQTQKSLMSAGIGFTIGSKTNTQENQSQS	600
20	orf117.pep	NEHTGSTVGS LKGD TTVAGKH YE QIGSTVSSPEGNN TIYAQSIDIOAAHNKLSNNTQT	210
	orf117ng	NEHTGSTVGS LKGD TTVASKHYEQ TGSNVSSPEGNN LISTQSM DIGAAQNLNSKTTQT	660
25	orf117.pep	YEQKXLT VAFSSPVTDLAQQ	230
	orf117ng	YEQKGLT VAFSSPVTDLAQQAIAVAHKA AKQFDKAKTTALMPWRLPMQVGR LFKQAKAPK	720

An ORF117ng nucleotide sequence <SEQ ID 519> was predicted to encode a protein having amino acid sequence <SEQ ID 520>:

25	1	..LLVQTEKDGL HNEQTFGEKK VFSENGKLHN YWRARRKGHD ETGHREQNYT
	51	LP EEITRDIS LGSFAYESH S KALSRHAPSQ GTELPQSNRD NIRTA KSNGI
	101	SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPFRFANYRQ WLGS DYMLGS
	151	LKLD PNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
	201	NGATAARSMN LSVGIALSAE QAAQLTSDIV WLQKEVKLP DGGTQTVLMP
30	251	QVYVRVKNGG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
	301	DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTL LLLNAGNN INNQSTAKSS
	351	QNAQGSSTYL DRMAGIYITG KEKGV LAAQA GKDINI IAGQ ISNQSDQGQT
	401	RLQAGR DINL DTVQTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDTVLL
	451	SGNNLNAAAEVGSAGKTL VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
35	501	GNKLVIDTKA QSHHETAQSS TFEGKQVVLQ AGNDANILGS NVISDNGTRI
	551	QAGNHVRIGT TQTQSQSEY HQTQKSLMS AGIGFTIGSK TNTQENQSQS
	601	NEHTGSTVGS LKGD TTVAS KHYEQTGSNV SSPEGNN LIS TQSM DIGAAQ
	651	NQLNSKTTQT YEQKGLT VAF SSPVTDLAQQ AIAVAHKA AK QFDKAKTTAL
	701	MPWRLPMQVG RLFKQAKAPK K*

40 Further work revealed the following gonococcal partial DNA sequence <SEQ ID 521>:

	1	TTGCTTGTGC AAACAGAAAA AGACGGTTTG CATAACGAGC AAACCTTTTG
	51	CGAGAAGAAA GTCTTCAGCG AAAATGGTAA GTTGCAACAAC TACTGGCGTG
	101	CGCGTCGTAA AGGACATGAT GAAACAGGGC ATCGTGAACA AAATTATACT
	151	TTGCCCGAGG AAATCACACG CGACATTTCA CTGGGTTTCA TTGCCTATGA
45	201	ATCGCATAGC AAAGCATTAA GCCGTCATGC GCCCAGCCAA GGCAC TGAGT
	251	TGCCACAAAG TAACCGGGAT AATATCCGTA CTGCGAAAAG CAACGGTATT
	301	TCGCTACCCT ATACGCCCAA TTCTTTTACC CCATTACCCG GCAGCAGCTT
	351	ATACATTATC AATCCTGCCA ATAAAGGCTA TCTTGTGTA ACCGATCCAC
	401	GCTTTGCCAA CTACCGTCAA TGGTTGGGTA GTGACTATAT GCTGGGCAGC
50	451	CTCAAACTAG ACCCAAACAA TTTACATAAA CGTTTGGGTG ATGGTTATTA
	501	CGAGCAACGT TTAATCAATG AACAAATCGC AGAGCTGACA GGGCATCGTC
	551	GTTTAGACGG TTATCAAAAC GACGAAGAAC AATTTAAAGC CTTAATGGAT
	601	AATGGCGCGA CTGCGGCACG TTCGATGAAT CTCAGCGTTG GCATTGCATT
	651	AAGTGCCGAG CAAGCAGCGC AACTGACCAG CGATATGTGT TGGTTGGTAC
55	701	AAAAAGAAAT TAACTTCCT GATGGCGGCA CACAAACCGT ATTGATGCCA
	751	CAGGTTTATG TACGCGTTAA AAATGGCGGC ATAGACGGTA AAGGTGCATT
	801	GTGTGTCAGG AGCAATACAC AAATCAATGT TTCAGGCAGC CTGAAAACT
	851	CAGGCACGAT TGCAGGCGC AATGCGCTTA TTATCAATAC CGATACGCTA
	901	GACAATATCG GTGGGCGTAT TCATGCGCAA AAATCAGCGG TTACGGCCAC
60	951	ACAAGACATC AATAATATTG GCGGCATTCT TTCTGCCGAA CAGACATTAT
	1001	TGCTCAATGC GGGTAACAAC ATCAACAACC AAAGCACGGC CAAGAGCAGT
	1051	CAAAATGCAC AAGGTAGCAG CACCTACCTA GACCGAATGG CAGGTATTTA

5
 10
 15
 20

```

1101 TATCACAGGC AAAGAAAAAG GTGTTTTAGC AGCGCAGGCA GGCAAAGACA
1151 TCAACATCAT TGCCGGTCAA ATCAGCAATC AATCAGATCA AGGGCAAACC
1201 CGGCTGCAGG CAGGACGCGA CATTAACTG GATACGGTAC AAACCGGCAA
1251 ATATCAAGAA ATCCATTTTG ATGCCGATAA CCATACCATC CGAGGTTCAA
1301 CGAACGAAGT CGGCAGCAGC ATCAAACAA AAGGCGATGT TACCCtatTG
1351 TCAGGGAATA ATCTCAATGC CAAAGCTGCC GAAGTCGGCA GCGCAAAGG
1401 CACACTTGCC GTGTATGCTA AAAATGACAT TACTATCAGC TCAGGCATCC
1451 ATGCCGGCCA AGTTGATGAT GCGTCCAAAC ATACAGGCAG AAGCGGCGGC
1501 GGTAATAAAT TAGTCATTAC CGATAAAGCC CAAAGTCATC ACGAAACTGC
1551 TCAAAGCAGC ACCTTTGAAG GCAAGCAAGT TGTATTGCAG GCAGGAAACG
1601 ATGCCAACAT CCTTGGCAGT AATGTTATT CCGATAATGG CACCCGGATT
1651 CAAGCAGGCA ATCATGTTTC CATTGGTACA ACCCAAACCTC AAAGCCAAAG
1701 CGAAACCTAT CATCAAACCC AAAAATCAGG ATTGATGAGT GCAGGTATCG
1751 GCTTCACTAT TGGCAGCAAG ACAACACAC AAGAAAACCA ATCCCAAAGC
1801 AACGAACATA CAGGCAGTAC CGTAGGCAGC CTGAAAGGCG ATACCACCAT
1851 TGTGCAAGC AAACACTACG AACAAACCGG CAGCAACGTT TCCAGCCCTG
1901 AGGGCAACAA CCTTATCAGC ACGCAAAGTA TGGATATTGG CGCAGCACAA
1951 AACCAATTAA ACAGCAAAAC CACCAAACCTC TACGAACAAA AAGGCTTAAC
2001 GGTGGCATTG AGTTGCGCCG TTACCGATT GGCACAACAA GCGATTGCCG
2051 TAGCACACAA AGCAGCAAAC AAGTCGGACA AAGCAAAAAC GACCGCGTTA
2101 ATGCCATGGC GGCTGCCAAT GCAGGTTGGC AGGCCTATCA AACAGGCAAA
2151 GCGCACAAA ACTTAG
  
```

This corresponds to the amino acid sequence <SEQ ID 522; ORF117ng-1>:

25
 30
 35

```

1 LLVQTEKDGL HNEQTFGEKK VSENGKLNH YWRARRKGHD ETGHREQNYT
51 LPEEITRDIS LGSFAYESHS KALSRHAPSQ GTELPQSNRD NIRTAKSNGI
101 SLPYTPNSFT PLPGSSLYII NPANKGYLVE TDPRFANYRQ WLGS DYMLGS
151 LKLDPNNLHK RLGDGYEQR LINEQIAELT GHRRLDGYQN DEEQFKALMD
201 NGATAARSMN LSVGIALSAE QAAQLTSDIV WLVOKEVKLP DGGTQTVLMP
251 QVYVRVKNNG IDGKGALLSG SNTQINVSGS LKNSGTIAGR NALIINTDTL
301 DNIGGRIHAQ KSAVTATQDI NNIGGILSAE QTLNAGN INNQSTAKSS
351 QNAQGSSTYL DRMAGIYITG KEKGVLAQA GKDINIAGQ ISNQSDQQT
401 RLQAGRDIHL DTVOTGKYQE IHFDADNHTI RGSTNEVGSS IQTKGDVTL
451 SGNLNAKAA EVGSAKGTLA VYAKNDITIS SGIHAGQVDD ASKHTGRSGG
501 GNKLVIDKA QSHHETAQSS TFEKGQVVLQ AGNDANILGS NVISDNQTRI
551 QAGNHVRIGT TQTSQSEY HQTQKSLMS AGIGFTIGSK TNTQENQSQS
601 NEHTGSTVGS LKGDTTIVAS KHYEQTGSNV SSPEGNNLIS TQSM DIGAAQ
651 NQLNSKTTQT YEOKGLTVAF SSPVTDLAQQ AIAVAHKAAN KSDKAKTTAL
701 MPWRLPMQVG RPIKQAKAHK T*
  
```

40 ORF117ng-1 shows the same 90% identity over a 230aa overlap with ORF117. In addition, it shows homology with a secreted *N.meningitidis* protein in the database:

45
 50
 55
 60
 65

```

gi|2623258 (AF030941) putative secreted protein [Neisseria meningitidis]Length =
2273
Score = 604 bits (1541), Expect = e-172
Identities = 325/678 (47%), Positives = 449/678 (65%), Gaps = 22/678 (3%)

Query: 1 LLVQTEKDGLHNEQTFGEKKVSENGKLNHYWRARRKGHD E TGHREQNYTLPEEITRDIS 60
L+V T + L N++T G K + ++ G L H Y R +KG D TG+ Y E++ I
Sbjct: 739 LIVGTPESALDND ETLGKTI-TDKGDLHRYHRHKKGRDSTGYSRSPYEPAPEVS-SIR 796

Query: 61 LGSFAYESHKALSRHAPSQGT ELPQSNRDNIRTAKSNGISLPYTPNSFTPLPGSSLYII 120
+G AY+ + AP Q +++P + + NGI +T LP SSL+ I
Sbjct: 797 MGISAYKGY-----APQQASDIPGTV---VPVVAENGIHPTFT-----LPNSSLFAI 840

Query: 121 NPANKGYLVETDPRFANYRQWLGS DYMLGSLKLDPNNLHKRLGDGYEQRLINEQIAELT 180
P NKGYL+ETDP F +YR+WLGS YML +L+ DPN++HKRLGDGYEQL+NEQIA+LT
Sbjct: 841 APNNKGYLIETDPAFTDYRKWLGS GYMLAALQDPNHIHKLRLGDGYEQLVNEQIAKL 900

Query: 181 GHRRLDGYONDEEQFKALMDNGATAARSMNLSVGIALSAEQAAQLTSDIVWLVOKEVKLP 240
G+RRLDGY NDEEQFKALMDNG T A+ + L+ GIALSAEQ A+LTSDIVWL + V LP
Sbjct: 901 GYRRLDGYTNDEEQFKALMDNGITIAKELQTPGIALSAEQVARLTSDIVWLENETVTLP 960

Query: 241 DGGTQTVLMPQVYVRVKNNGIDGKGALLSGSNTQINVSGSLKN-SGTIAGRNALIINTDT 299
DG TQTVL P+VYVR + ++G+ALLSGS I SG+++N G IAGR ALI+N
Sbjct: 961 DGTQTVLKPQVYVRARPKMDNGQALLSGSVVDIG-SGAIENRGGLIAGREALILNAQN 1019

Query: 300 LDNIGGRIHAQKSAVTATQDINNIGGILSAEQTLNAGNINNINQSTAKSSQNAQGSSTY 359
  
```

-307-

+ N+ G + + A D I N G I A E L L L A N N I ++S +S+QN QGS
 Sbjct: 1020 IKNLQGDLOGKNIFAAAGSDITNTGSI-GAENALLLKASNNIESRSETRSNQNEQGSVRN 1078
 Query: 360 LDRMAGIYITGKEKGVLAAGKGDINIIAGQISNOSDQGGQTRLQAGRDINLDTVQTGKYQ 419
 + R+AGIY+TG++ G + AG +I + A +++NQS+ GQT L AG DI DT + Q
 Sbjct: 1079 IGRVAGIYLTGRQNGSVLLDAGNNIVLTASELTNQSEDGQTVLNAGGDIRSCTTGISRNO 1138
 Query: 420 EIHFDADNHTIRGSTNEVGSSIQTKGDVTLISGNNLNAAAEVGSAGKTLAVYAKNDITI 479
 FD+DN+ IR NEVGS+I+T+G+++L + ++ +AAEVGS +G L + A DI +
 Sbjct: 1139 NTIFDSNYVIRKEQNEVGSTIRTRGNLSLNAKGDIRIRAAEVGSEQRLKLAAGRDIKV 1198
 Query: 480 SSGIHAGQVDDASKHTGRSGGGNKLVTDKAQSHHETAQSSSTFEGKQVVLQAGNDANILG 539
 +G + +DA K+TGRSGGG K +T ++ + A S T +GK+++L +G D + G
 Sbjct: 1199 EAGKAHTETEDALKYTGRSGGGIKQKMRHLKNQNGQAVSGTLDGKEIILVSGRDITVTG 1258
 Query: 540 SNVISDNTRIQAAGNHVRIGTTQTQSQSEYHQTQKSGLM-SAGIGFTIGSKTNTQENQS 598
 SN+I+DN T + A N++ + +T+S+S ++ +KSGLM S GIGFT GSK +TQ N+S
 Sbjct: 1259 SNIIADNHTILSAKNIVLKAETRSRSAEMNKKEKSGLMGSGGIGFTAGSKKDTQTNR 1318
 Query: 599 QSNEHTGSTVGSILKGDTTIVASKHYEQTGSNVSSPEGNNLISTQSMIDIGAAQNQLNSKTT 658
 ++ HT S VGSL G+T I A KHY QTGS +SSP+G+ IS+ + I AAQN+ + ++
 Sbjct: 1319 ETVSHTESVVGSLNGNTLISAGKHYTQTGSTISSPQGDVGISSGKISIDAAQNRYSQESK 1378
 Query: 659 QTYEQKGLTVAFSSPVTD 676
 Q YEQKG+TVA S PV +
 Sbjct: 1379 QVYEQKGVTVAVISVPVVN 1396

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 63

30 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 523>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTCGCCTA
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCTG
 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAWAACCAG CCATGTCCGC
 151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCAACCGGC
 201 GGTCAAAAAA ACGGCAAAAC CCAAGACCC CGYCATGCGC AACCTGCAAG
 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
 301 TTCAAAACCG AAATCGAAAC CGCCTTGAA GAAAGCGGCA TTATCGGCAA
 351 CTCCGCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCAACGAAAC
 401 CTGCCGACGC GTCGGCAAAA CCTGCACCCG TTCCGCAAAAC ACCTGCAAAA
 451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAT CCTGGTTTGA
 501 CGTGCATCATC GACTTCATCT CCTAT...

This corresponds to the amino acid sequence <SEQ ID 524; ORF119>:

1 MIYIVLFLAV VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSXTSHVR
 51 DGKPSGGSSVM MPKPQPAVKK TAKPQDPXMR NLQEQDAVYI AKQKQAKASP
 101 FKTEIETALE ESGIIGNSAH TVSEPQTGHS ATKPADASAK PAPVPQTPAK
 151 PLITLKELSK VELSWFDVRI DFISY...

Further work revealed the complete nucleotide sequence <SEQ ID 525>:

1 ATGATTTACA TCGTACTGTT TCTAGCTGTC GTCCTCGCCG TTGTCGCCTA
 51 CAACATGTAT CAGGAAAACC AATACCGCAA AAAAGTGCGC GACCAGTTCTG
 101 GACACTCCGA CAAAGATGCC CTGCTCAACA GCAAAACCAG CCATGTCCGC
 151 GACGGCAAAC CGTCCGGCGG GTCAGTCATG ATGCCGAAAC CCAACCGGC
 201 GGTCAAAAAA ACGGCAAAAC CCAAGACCC CGCCATGCGC AACCTGCAAG
 251 AACAGGATGC CGTCTACATC GCCAAGCAGA AACAGGCAAA AGCCTCCCCG
 301 TTCAAAACCG AAATCGAAAC CGCCTTGAA GAAAGCGGCA TTATCGGCAA
 351 CTCCGCCAC ACCGTTTCCG AACCCCAAAC CGGACATTCC GCAACGAAAC
 401 CTGCCGACGC GTCGGCAAAA CCTGCACCCG TTCCGCAAAAC ACCTGCAAAA
 451 CCGCTGATTA CGCTCAAAGA ACTGTCAAAA GTCGAATTAC CCTGGTTTGA
 501 CGTGCCTTTC GACTTCATCT CCTATATCGC GCTGACCGAA GCCAAAGAAC
 551 TGCACGCACT GCGCGCCCTT TCCAACCGCT GCGGTACCA GATTGTCGCG
 601 TGCACCATGG ACGACATTT CCAGATTGCC GAACCATCC CGGGCATCCG

10

651	CTATCAGGCA	TTTATCGTGG	GTATTCAGGC	AGTCAGCCGC	AACGGACTTG
701	CCTCGCAGGA	AGAACTCTCC	GCATTCAACC	GCCAGGTGGA	CGCATTTCGA
751	CAAAGCATGG	GCGCTCAGAC	GCTGCACACC	GACCTTGCCG	CCTTTATCGA
801	AGTGGCTTCC	GCGCTGACGC	CATTCTGCGC	GCGCGTCGAC	CAGACCATCG
851	CCATCCATTT	GGTTTCCCGC	ACCAGCATCA	GCGGCGTAGA	ATCGCGTTCC
901	GCGGTAACGG	GCGTGGGTTT	CGTTTTGGAA	GACGACGGCG	CGTTCCTACT
951	TACCGACACG	TCGGGCTCGA	CCATTGTCTC	CATCTGCTCG	CTCAACAACG
1001	ATGCGGTTTAC	CAACGCCCTT	TTGGACAACC	AGTCCTACAA	AGGCTTCAGT
1051	ATGCTGCTCG	ACATCTCCGCA	CTCTCCGGCA	GGGGAATAAA	CCTTCGACGA
1101	TTTGTTTATG	GATTTGGCGG	TACGCGTGTC	CGGCCAGTTG	AACCTGAATC
1151	TGGTCAACGA	CAAAATGGAA	GAAGTTTCGA	CCCAATGGCT	CAAAGACGTG
1201	CGCATTATG	TATTGGCGCG	TCAGTCCGAG	ATGCTCAAAG	TCGGTATCGA
1251	ACCGGGCGGC	AAAACCGCAT	TGCGCCTGTT	CTCCTAA	

15

20

```

1  MIYIVLFLAV  VLAVVAYNMY  QENQYRKKVR  DQFGHSDKDA  LLNSKTSHVR
51  DGKPSGGSSVM  MPKPQPAVKK  TAKPQDPAMR  NLQEQDAVYI  AKOKQAKAPF
101 FKTEIETALE  ESGIIGNSAH  TVSEPQTGHS  APKPADAPAK  PAPVPQTPAK
151 PLITLKELSK  VELPWFVDFR  DFISYIALTE  AKELHALPRL  SNCRCYQIVG
201 CTMDHDFQIA  EPIPIGRYQA  FIVIGQAVSR  NGLASQEELS  AFNRQVDafa
251 QSMGGQTLHT  DLAAFIEVAS  ALDAFCARVD  QTIAIHLVSP  TSIISGVELRS
301 AVTGVGVFLE  DDGAFHYTDT  SGSTMFISCS  LNEPFPTNAL  LDNQSYKGFS
351 MLLDIPHSEA  GEKTFDDLTD  DLAVRLSGQL  NLNLVNDKME  EVSTQWLKDV
401 RTYVLAROSE  MLKVGIPEGG  KTLARLFS*

```

25

ORF119 shows 93.7% identity over a 175aa overlap with an ORF (ORF119a) from strain A of *N.*

meningitidis:

30

35

40

45

	10	20	30	40	50	60
orf119.pep	MIYIVLFLAVVLAVVAYNMYQENQYRKKVDRQFGHSDKDALLNSXTS	SHVRD	GKPS	GGSV	M	
	:					
orf119a	MIYIVLFLAAVLAVVAYNMYQENQYRKKVDRQFGHSDKDALLNSKTS	SHVRD	GKPS	GGPVM		
	10	20	30	40	50	60
	70	80	90	100	110	120
orf119.pep	MPKPQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH					
orf119a	MPKPQPAVKKTAKSQDPAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEESGIIGNSAH					
	70	80	90	100	110	120
	130	140	150	160	170	
orf119.pep	TVSEPQTGHSATKPADASAKPAPVPQT	PAKPLITL	KELSKVELSWFD	VRD	FISY	
orf119a	TVPEPQTGHSAPKPADAPAKPVPVPQT	PAKPLITL	KELSKVELPWFD	VRD	FISYIALTE	
	130	140	150	160	170	180
orf119a	AKELHALPRLSNRCRYQIVGCTMDDHFQIAEPIPGIRYQAFIVGIQAVSRNGLASQEELS					
	190	200	210	220	230	240

50

55

60

1	ATGATTTACA	TCGTACTGTT	CCTCGCCGCC	GTCTCGCG	TTGTGCGCTA
51	CAATATGTAT	CAGGAAAACC	AATACCGCAA	AAAAGTCGCG	GACCAAGTTCG
101	GGCACTCCGA	CAAAAGATGCC	CTGCTCAACA	GCAAAACCAAG	CCATGTCGCG
151	GACGGCAAC	CGTCGGCGG	GCCAGTCATG	ATGCGGAAAC	CCCAACCCGGC
201	GGTCAAAAAA	ACGGCAAAAT	CCCAAGACCC	CGCCATGCGC	AACCTGCAAG
251	AGCAGGATGC	CGTCTACATC	GCCAAGCAGA	AACAGGCAAA	AGCCTCCCGC
301	TTCAAAACCG	AAATCGAAAC	CGCCTTGGA	GAAGCGGCA	TTATCGGCAA
351	CTCCGCCAC	ACCGTCCCG	AACCCCAAC	CGGACATTCC	GACCAACAA
401	CTGCCGACG	CTGGTCCAAA	CTGTTTCCG	TTCCGCAAC	GCCGGCAAAA
451	CCGCTGATTA	CGCTCAAAGA	GCTGTGGAAG	GTGAGCTGCG	CCTGGTTTGA
501	CGTGCGCTTC	GACTTCATCT	CTTATATCGC	GCTGACCGAA	GCCAAAGAAG
551	TGCACGCACT	CGCGCGCTT	TCCAACCGCT	CGCGTACCA	GATTGTGCGC
601	TGCACCATGG	ACGACCATTT	CCAGATTGCC	GAACCCATCC	CGGGCATCCG

5 651 CTATCAGGCA TTTATCGTGG GTATTTCAGGC AGTCAGCCGC AACGGACTTG
 701 CCTCGCAGGA AGAACTCTCC GCATTCAACC GCCAGGTGGA TGCATTTCGCA
 751 CACAGCATGG GCGGTTCAGAC GCTGCACACC GACCTTGCCG CCTTTATCGA
 801 AGTGGCTTCC GCACTGGACG CATTCTGCGC GCGCGTCGAC CAGACTATCG
 851 CCATCCATTT GGTTCCTCCG ACCAGCATCA GCGCGCTAGA ACTGCGTTCC
 901 GCCGTAACGG GCGTGGGTTT CGTTTTGGAA GACGACGGCG CGTTCCTACTA
 951 TACCGACACG TCGGGCTCGA CCATGTTCTC CATCTGCTCG CTCAACAACG
 1001 AGCCGTTTAC CAATGCCCTT TTGGACAACC AGTCTATAA AGGCTTCAGT
 1051 ATGCTGCTCG ACATCCCGCA CTCTCCGGCA GGCGAAAAA CCTTCGACGA
 1101 TTGTTTATG GATTTGGCGG TACGCCTGTC CGGCCAGTTG AACCTGAATC
 1151 TGGTCAACGA CAAATGGAA GAAGTTTCGA CCAATGGCT CAAAGACGTG
 1201 CGCACTTATG TATTGGCTCG TCAGTCCGAG ATGCTCAAAG TCGGTATCGA
 1251 ACCGGGCGGC AAAACCGCAT TCGCCTGTT CTCCTAA

This encodes a protein having amino acid sequence <SEQ ID 528>:

15 1 MIYIVLFLAA VLAVVAYNMY QENQYRKKVR DQFGHSDKDA LLNSKTSHVR
 51 DGKPSGGPVM MPKPQPAVKK TAKSQDPAMR NLQEQDAVYI AKQKQAKASP
 101 FKTEIETALE ESGIIGNSAH TVPEPQTGHS APKPADAPAK PVPVPQTPAK
 151 PLITLKELSK VELPWFDFVR DFISYIALTE AKELHALPRL SNRCRYQIVG
 201 CTMDDHFQIA EPIPGIRYQA FIVGIQAVSR NGLASQEELS AFNRQVDAFA
 251 HSMGGQTLHT DLAAFIEVAS ALDAFCARVD QTIHILVSP TSISGVELRS
 301 AVTGVGFVLE DDGAFHYTDT SGSTMFSICS LNNEPFTNAL LDNQSYKGFS
 351 MLLDIPHSPA GEKTFDDLFM DLAVRLSGQL NLNLVNDKME EVSTQWLKDV
 401 RTYVLARQSE MLKVGIEPGG KTALELRF*

ORF119a and ORF119-1 show 98.6% identity in 428 aa overlap:

25	orf119a.pep	10	20	30	40	50	60
	orf119-1	10	20	30	40	50	60
30	orf119a.pep	70	80	90	100	110	120
	orf119-1	70	80	90	100	110	120
35	orf119a.pep	130	140	150	160	170	180
	orf119-1	130	140	150	160	170	180
40	orf119a.pep	190	200	210	220	230	240
	orf119-1	190	200	210	220	230	240
45	orf119a.pep	250	260	270	280	290	300
	orf119-1	250	260	270	280	290	300
50	orf119a.pep	310	320	330	340	350	360
	orf119-1	310	320	330	340	350	360
55	orf119a.pep	370	380	390	400	410	420
	orf119-1	370	380	390	400	410	420

5 Homology with a predicted ORF from *N.gonorrhoeae*

	orf119.pep	MIYIVLFLAVVLAVVAYNMYQENQYRKVKVRDQFGHSDKDALLNSXTSHVRDGGKPSGGSV	60
10	orf119ng	MIYIVLFLAAVLAVVAYNMYQENQYRKVKVRDQFGHSDKDALLNSKTSHVRDGGKPSGGPVM	60
	orf119.pep	MPKPQPAVKKTAKPQDPXMRNLQEQDAVYIAKQKQAKASPFKTEIETALESIGIINSAH	120
15	orf119ng	MPKPQPAVKKPAKPQDSAMRNLQEQDAVYIAKQKQAKASPFKTEIETALEEIGIINSAH	120
	orf119.pep	TVSEPQTGHSATKPADASAKPAPVPTPAKPLITLKELSKVELSWFDVRIDFISY	175
	orf119ng	TVSEPQTGHSAPKPADAPAKPVFPVPTPAKPLITLKELSKVELPWFDFRDFISYIALTE	180

20	1	ATGATTTACA	TCGTACTGTT	CCTCGCGGCC	GTCCTCGCGG	TTGTGCGCTA
	51	CAATATGTAT	CAGGAAAACC	AATACCGCAA	AAAAGTGCGC	GACCAGTTTCG
	101	GACACTCCGA	CAAGAGATGCC	CTGCTCAACA	GCAAAACCCAG	CCATGTCCCGC
	151	GACGGCAAA	CGTCCGGCGG	GCCAGTCATG	ATGCCGAAAC	CCCAACCGGC
	201	GGTCAAAAAA	CCGGCCAAAC	CCCAGACTC	CGCCATGCGC	AACCTGCAAG
25	251	AACAGGATGC	CGTCTACATC	GCCAAGCAGA	AACAGGCCAA	AGCCTCCCCG
	301	TTCAAAACCG	AAATCGAAAC	CGCCTTGGAA	GAATCTGGCA	TTATCGGCAA
	351	CTCCGCCCAC	ACCGTTTCCG	AACCCCAAAC	CGGACATTCC	GACCCGA AAC
	401	CTGCCGACGC	GCCGGCAAAA	CCC GTTCCCG	TTCCGCAAA	GCCGGCAAAA
	451	CCGCTGATTA	CGCTCAAGAA	GCTGTGCAAG	GTCGAGCTGC	CCTGGTTTGA
30	501	CGTGCGCTtc	gACTTCATCT	CCTATATCCG	GCTGACCGAA	GCCAAAGAAC
	551	TGCACGCACT	GCCGCGCCTT	tcAACCGCT	GCCGCTACCA	GATTGTCCGC
	601	TGCACCATGG	ACGACCATT	CCAGATTGCC	GAACCCATCC	CGGGCATCTG
	651	CTATCAGGCA	TTTATCGTGG	GTATCCAGG	AGTCAGCCGC	AACGGACTTG
	701	CCTCGCAGGA	AGAACTCTCC	GCATTCAACC	GCCAGCGCGA	CGGATTCCGA
35	751	CAAAGCATGG	GCGGTGAGAC	GCTGCACACC	GACCTTGCCG	CCTTTATCGA
	801	AGTGGCTTCC	GCATCTGGACG	CATTCTGCGC	GCGCGTGCAC	GACAGCATCG
	851	CCATCCATT	GGTTCGCGG	ACCAGCATCA	GCGCGGTAGA	ACTGCGTTTG
	901	GCGGTAAACGG	CGTGGGTTT	CGTTTTGGAA	GAGCAGCGCG	CGTTCCACTA
	951	TACCGACACG	TCGGGCTCGA	CCATGTTCTC	CATCTGCTCG	CTCAACAACG
40	1001	AGCCGTTTAC	CAATGCCCTT	TTGGACAACC	AGTCCTACAA	AGGCTTCAGT
	1051	ATGCTGCTCG	ACATCCGCGA	CTCTCCGGCA	GCGCAAAAAA	CCTTCGACGA
	1101	TTGTTTATG	GATTGTGGCG	TACGCCGTGC	CGGTCAAGTT	AACCTGAATC
	1151	TGGTCAACGA	CAAAATGGCA	GAAATTTTCA	CCCAATGGCT	CAAAGACGTA
	1201	GCGACTTATG	TATTGGCGCG	TCAGTCCGAG	ATGCTCAAAG	TCGGTATCGA
45	1251	ACCGGGCGGC	AAAACCGCCC	TGCGCCTGTT	TTCTATAA	

	1	MIYIVFLAA	VLAVVAYNMY	QENQYRKKVR	DQFGHSDKDA	LLNSKTSHVR
	51	DGKPSGGPVM	MPKPQPAVKK	PAKPQDSAMR	NLQEQDAVYI	AKQKQAKASP
50	101	FKTEIETALE	EIGIIGNSAH	TVSEPQTGHS	APKPADAPAK	PVPVPQTPAK
	151	PLITLKELSK	VELPWFVDVR	DFISYIALTE	AKELHALPRL	SNRCRYQIVG
	201	CTMDHDFQIA	EPIPGIRYQA	FIVGIIQAVSR	NGLASQEELS	AFNRQADAFa
	251	QSMGGQTLHT	DLAAFIEVAS	ALDAFCARVD	QTIATHLVSP	TSISGVELRS
	301	AVTGVGFVLE	DDGAFHYTDT	SGSTMFSICS	LNNPFTNAL	LDNQSYKGFS
	351	MLLDIPHSA	GAKTFDDLFM	DLAVRLSGQL	NLNLVNDKME	EVSTQWLKDV
55	401	RTYVLARQSE	MLKVTIEPFG	KTALRLFS*		

```

              10      20      30      40      50      60
orf119ng      MIYIVLFLAAVLAVVAYNMYQENQYRKKVRDQFGHSDKDALNSKTSHV RDGKPSGGFVM
              |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 orf119-1    MIYIVLFLAVVLAVVAYNMYQENQYRKKVRDQFGHSDKDALNSKTSHV RDGKPSGGFSVM
              10      20      30      40      50      60

```

-311-

		70	80	90	100	110	120
	orf119ng	MPKPQPAVKKPAKQDSAMRN	LQEQDAVYIAKQKQAKAS	PFKTEIETALEEIGIIGNSAH			
5	orf119-1	MPKPQPAVKKTAKPQDPAMRN	LQEQDAVYIAKQKQAKAS	PFKTEIETALEESGIIGNSAH			
		70	80	90	100	110	120
	orf119ng	TVSEPTGHSAPKPADAPAKPVP	PQTPAKPLITL	KELSKVELPWFDV	RDFISYIALTE		
10	orf119-1	TVSEPTGHSAPKPADAPAKPVP	PQTPAKPLITL	KELSKVELPWFDV	RDFISYIALTE		
		130	140	150	160	170	180
	orf119ng	AKELHALPRLSNRCRYQIVGCT	MDDHFQIAEPI	PGIRYQAFIVGIQAVSRNGLASQEELS			
15	orf119-1	AKELHALPRLSNRCRYQIVGCT	MDDHFQIAEPI	PGIRYQAFIVGIQAVSRNGLASQEELS			
		190	200	210	220	230	240
	orf119ng	AFNRQADAFASQMGQTILHTDL	AAFIEVASALDAFCARVDQTIAI	HLVSP	TSISGVELRS		
20	orf119-1	AFNRQVDAFAQSMGGQTILHTDL	AAFIEVASALDAFCARVDQTIAI	HLVSP	TSISGVELRS		
		250	260	270	280	290	300
	orf119ng	AVTGVGFVLEDDGAFHYTDTSG	TMFSICSLNNEPFTNALLDNQSYK	GFSMLLDIPHSPA			
25	orf119-1	AVTGVGFVLEDDGAFHYTDTSG	TMFSICSLNNEPFTNALLDNQSYK	GFSMLLDIPHSPA			
		310	320	330	340	350	360
	orf119ng	GEKTFDDLFMDLAVRLSGQLNL	NLVNDKMEEVSTQWLKDVRTYV	LARQSEMLKVGIEP	GG		
30	orf119-1	GEKTFDDLFMDLAVRLSGQLNL	NLVNDKMEEVSTQWLKDVRTYV	LARQSEMLKVGIEP	GG		
		370	380	390	400	410	420
	orf119ng	KTALRLFSX					
35	orf119-1	KTALRLFSX					
		429					
40	orf119ng	KTALRLFSX					
	orf119-1	KTALRLFSX					

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

45 Example 64

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 531>

	1	..GCGCGGCACG	GCACGGAAGA	TTTCTTCATG	AACAACAGCG	ACAC.ATCAG
	51	GCAGATAGTC	GAAAGCACCA	CCGGTACGAT	GAAGCTGCTG	ATTCCTCCA
	101	TCGCCCTGAT	TTCATTGGTA	GTCGGCGGCA	TCGGCGTGAT	GAACATCATG
50	151	CTGGTGTCG	TTACCGAGCG	CACCAAAGAA	ATCGGCATAC	GGATGGCAAT
	201	CGGCGCGCGG	CGCGGCAATA	TTTyGCAGCA	GTTTTTGATT	GAGGCGGTGT
	251	TAATCTGCGT	CATCGGCGGT	TTGGTCGGCG	TGGGTTTGTC	CGCCGCCGTC
	301	AGCCTCGTGT	TCAATCATTT	TGTAACCGAC	TTCCCGATGG	ACATTCCGC
	351	CATGTCCGTC	ATCGGCGCGG	TCGCCTGTTC	GACCGGAATC	GGCATCGCGT
55	401	TCGGCTTTAT	GCCTGCCAAT	AAAGCAGCCA	AACTCAATCC	GATAGACGCA
	451	TTGGCACAGG	ATTGA			

This corresponds to the amino acid sequence <SEQ ID 532; ORF134>:

	1	..ARHGTEDEFFM	NNSDXIRQIV	ESTTGTMKLL	ISSIALISLV	VGGIGVMNIM
	51	LVSVTERKE	IGIRMAIGAR	RGNIXQQFLI	EAVLICVIGG	LVGVGLSAAV
60	101	SLVFNHFVTD	FPMDISAMSV	IGAVACSTGI	GIAFGFMPAN	KAAKLNPIDA
	151	LAQD*				

Further work revealed the complete nucleotide sequence <SEQ ID 533>:

```

      1 ATGTCGGTGC AAGCAGTATT GGCGCACAAA ATGCGTTCGC TTCTGACGAT
      51 GCTCGGCATC ATCATCGGTA TCGCGTCGGT GGTTTCCGTC GTCCGATTGG
    101 GCAATGGTTC GCAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
    151 AACACCATCA GCATCTTCCC GGGGCGCGGC TTCGGCGACA GGCGCAGCGG
    201 CAGGATTAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
    251 GCTACGTTGC TTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACT
    301 TACCGCAACA CCGACCTGAC CGCCTCGCTT TACGGCGTGG GCGAACAATA
    351 TTTCGACGTG CGCGGACTGA AGCTGGAAC GGGGCGGCTG TTTGACGAAA
    401 ACATGTGAA AGAAGACGCG CAGGTCGTCG TCATCGACCA AAATGTCAAA
    451 GACAACTCT TTGCGGACTC GGATCCGTTG GGTAAAACCA TTTTGTTCAG
    501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAGAC GAAAAACGCTT
    551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGCTGATG
    601 CACCAAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
    651 AGACAATGCC AATACCCAGG TTGCCGAAAA AGGGCTGACC GATCTGCTCA
    701 AAGCGCGGCA CGGCACGGA GATTCTTCA TGAACAACAG CGACAGCATC
    751 AGGCAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTCTCTC
    801 CATCGCCCTG ATTTTCATTG TAGTCGGCGG CATCGGCGTG ATGAACATCA
    851 TGCTGGTGTC CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
    901 ATCGGCGCGC GGGCGCGCAA TATTTTGAG CAGTTTTTGA TTGAGGCGGT
    951 GTTAATCTGC GTCATCGGCG GTTGGTTCGG CGTGGGTTTG TCCGCCGCCG
   1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCCGAT GGACATTTC
   1051 GCCATGTCCG TCATCGGCGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
   1101 GTTCGGCTTT ATGCCTGCCA ATAAAGCAGC CAACTCAAT CCGATAGACG
   1151 CATTGGCACA GGATTGA
  
```

This corresponds to the amino acid sequence <SEQ ID 534; ORF134-1>:

```

      1 MSVQAVLAHK MRSLLTMLGI IIGIASVSV VALNGSQKK ILEDISSIGT
      51 NTISIFPGRG FGDRRSGRIK TLTIDDAKII AKQSYVASAT PMTSSGGTIL
    101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
    151 DKLFDSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
    201 HQITGESHTN SITVKIKDNA NTQVAEKGLT DLLKARHGTE DFFMNSDSI
    251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
    301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFPMDIS
    351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*
  
```

35 Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o648 of *E. coli* (accession number AE000189)

ORF134 and o648 protein show 45% aa identity in 153aa overlap:

```

Orf134: 2  RHGTEDFFMNSDXIRQIVESTTGTMKXXXXXXXXXXXXXVVGIGVMNIMLVSVTERTKEI 61
          RHG +DFF N D + + VE TT T++ VVGIGVMNIMLVSVTERT+EI
    40 o648: 496 RHGKKDFFTWNMDGVLTVEKTRTLQLFLTLVAVISLVVVGIGVMNIMLVSVTERTREI 555

Orf134: 62  GIRMAIGARRGNIXQQFLIEAXXXXXXXXXXXXXXXXXXFNHEVTDFFPMDISAMSVI 121
          GIRMA+GAR ++ QQFLIEA F+ + + S ++++
    45 o648: 556 GIRMAVGARASDVLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPALL 615

Orf134: 122 GAVACSTGIGIAFGFMPANKAAKLNPIDALAQD 154
          A CST GI FG++PA AA+L+P+DALA++
    45 o648: 616 LAFLCSTVTGILFGWLPARNAARLDPVDALARE 648
  
```

50 Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF134 shows 98.7% identity over a 154aa overlap with an ORF (ORF134a) from strain A of *N.*

meningitidis:

```

      10      20      30
    55 orf134.pep      ARHGTEDFFMNSDXIRQIVESTTGTMKLL
          |||||
    55 orf134a      GESHTNSITVKIKDNANTQVAEKGLTDLKARHGTEDFFMNSDSIRQIVESTTGTMKLL
          210      220      230      240      250      260
          40      50      60      70      80      90
  
```

-313-

```

5  orf134.pep  ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG
   orf134a    ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICVIGG
                                270      280      290      300      310      320

10  orf134.pep  LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
   orf134a    LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA
                                330      340      350      360      370      380

15  orf134.pep  LAQDX
   orf134a    LAQDX

```

The complete length ORF134a nucleotide sequence <SEQ ID 535> is:

```

1  ATGTCGGTGC AAGCAGTATT GGCGCACAAA ATGCGTTCGC TTCTGACGAT
51  GCTCGGCATC ATCATCGGTA TCGCTTCGGT TGTCTCCGTC GTCGCATTGG
101 GCAACGGTTC GCAGAAAAAA ATCCTTGAAG ACATCAGTTC GATAGGGACG
151 AACACCATCA GCATCTTCCC AGGGCGCGGC TTCGGCGACA GGCGCAGCGG
201 CAGGATTAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
251 GCTACGTTGC TTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACT
301 TACCGCAATA CCGACCTGAC CGCTTCTTTG TACGGTGTGG GCGAACAATA
351 TTTCCAGCTG CGCGGGCTGA AGCTGGAAC GGGCGGGCTG TTTGACGAAA
25  401 ACGATGTGAA AGAAGACGCG CAGGTCGTCG TCATCGACCA AAATGTCAAA
451 GACAACTCTT TTGCGGACTC GGATCCGTTG GGTAAAACCA TTTTGTTCAG
501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAGAC GAAAAACGCTT
551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGGTGATG
30  601 CACCAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAAATCAA
651 AGACAAATGCC AATACCCAGG TTGCCGAAAA AGGGCTGACC GATCTGCTCA
701 AAGCGCGGCA CGGCACGGAA GATTCTTCA TGAACAACAG CGACAGCATC
751 AGGCAGATAG TCGAAAGCAC CACCGGTACG ATGAAGCTGC TGATTTCCTC
801 CATCGCCCTG ATTTATTGG TAGTCGGCGG CATCGGCGTG ATGAACATCA
851 TGCTGGTGTC CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
35  901 ATCGGCGCGC GCGCGGCAA TATTTGCAG CAGTTTGA TTAGGCGGT
951 GTTAATCTGC GTCATCGGCG GTTGGTTCG CGTGGGTTG TCCGCCGCCG
1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ACTTCCCGAT GGACATTTC
1051 GCCATGTCCG TCATCGGCGC GGTGCGCTGT TCGACCGGAA TCGGCATCGC
1101 GTTCGGCTTT ATGCTGCCA ATAAAGCAG CAAACTCAAT CCGATAGATG
40  1151 CATTGGCGCA GGATTGA

```

This encodes a protein having amino acid sequence <SEQ ID 536>:

```

1  MSVQAVLAHK MRSLLTMLGI IIGIASVSV VALNGSQKK ILEDISSIGT
51  NTISIFPGRG FGDRRSGRIK TLTIDDAKII AKQSYVASAT PMTSSGGTLT
101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
151 DKLFADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
45  201 HQITGESHTN SITVKIKDNA NTQVAEKGLT DLLKARHGTE DFFMNSDSI
251 RQIVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
301 IGARRGNILQ QFLIEAVLIC VIGGLVGVGL SAAVSLVFNH FVTDFFPMDIS
351 AMSVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

```

50 ORF134a and ORF134-1 show 100.0% identity in 388 aa overlap:

```

55  orf134a.pep  MSVQAVLAHKMRSLLTMLGIIIGIASVSVVALNGSQKKILEDISSIGTNTISIFPGRG
   orf134-1      MSVQAVLAHKMRSLLTMLGIIIGIASVSVVALNGSQKKILEDISSIGTNTISIFPGRG

   orf134a.pep  FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
   orf134-1      FGDRRSGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV

60  orf134a.pep  RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD
   orf134-1      RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTILFRKRPLTVIGVMKKD

   orf134a.pep  ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE
65  orf134-1      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLLKARHGTE

```


5 orf134a.pep DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
 orf134-1 DFFMNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
 orf134a.pep IGARRGNILQQFLIEAVLICVIGGLVGVLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC
 orf134-1 IGARRGNILQQFLIEAVLICVIGGLVGVLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC
 10 orf134a.pep STGIGIAFGFMPANKAAKLNPIDALAQDX
 orf134-1 STGIGIAFGFMPANKAAKLNPIDALAQDX

Homology with a predicted ORF from *N.gonorrhoeae*

15 ORF134 shows 96.8% identity over a 154aa overlap with a predicted ORF (ORF134.ng) from *N. gonorrhoeae*:

orf134.pep ARHGTEFFMNSDXIRQIVESTTGTMKLL 30
 orf134ng GESHTNSITVKIKDNANTRVAEKGLELLKARHGTEFFMNSDSIRQMVESTTGTMKLL 264
 20 orf134.pep ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNIXQQFLIEAVLICVIGG 90
 orf134ng ISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMAIGARRGNILQQFLIEAVLICIGG 324
 25 orf134.pep LVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVACSTGIGIAFGFMPANKAAKLNPIDA 150
 orf134ng LVGVGLSAAVSLVFNHFVTDFFPMDISAAVIGAVACSTGIGIAFGFMPANKAAKLNPIDA 384
 30 orf134.pep LAQD 154
 orf134ng LAQD 388

The complete length ORF134ng nucleotide sequence <SEQ ID 537> is:

1 ATGTCGGTGC AAGCAGTATT GCGGCACAAA ATGCGTTCGC TTCTGACCAT
 51 GCTCGGCATC ATCATCGGTA TCGCTTCGGT TGTCTCCGTC GTCGCGCTGG
 35 101 GCAACGGTTC GCAGAAAAAA ATCCTCGAAG ACATCAGTTC GATGGGGACG
 151 AACACCATCA GCATCTTCCC CGGGCGCGGC TTCGGCGACA GGCGCAGCGG
 201 CAAAATCAAA ACCCTGACCA TAGACGACGC AAAAATCATC GCCAAACAAA
 251 GCTACGTTGC CTCCGCCACG CCCATGACTT CGAGCGGCGG CACGCTGACC
 301 TACCGCAATA CCGACCTGAC CGCTTCTTTG TACGGTGTGG CGGAACAATA
 40 351 TTTCGACGTG CGCGGGCTGA AGCTGGAAC GGGGCGGCTG TTTGATGAGA
 401 ACGATGTGAA AGAAGACGCG CAAGTCGTCG TCATCGACCA AAATGTCAAA
 451 GACAACTCT TTGCGGACTC GGATCCGTTG GGTAAAACCA TTTTGTTCAG
 501 GAAACGCCCC TTGACCGTCA TCGGCGTGAT GAAAAAGAC GAAAACGCTT
 551 TCGGCAATTC CGACGTGCTG ATGCTTTGGT CGCCCTATAC GACGCTGATG
 45 601 CACCAAATCA CAGGCGAGAG CCACACCAAC TCCATCACCG TCAAATCAA
 651 AGACAATGCC AATACCCGGG TTGCCGAAA AGGGCTGGCC GAGCTGCTCA
 701 AAGCACGGCA CGGCACGGAA GACTTCTTTA TGAACAACAG CGACAGCATC
 751 AGGCAGATGG TCGAAAGCAC CACCGTACG ATGAAGCTGC TGATTTCCTC
 801 CATCGCCCTG ATTTTCATTG TAGTCGGCGG CATCGGTGTG ATGAACATTA
 50 851 TGCTGGTGTG CGTTACCGAG CGCACCAAAG AAATCGGCAT ACGGATGGCA
 901 ATCGGCGCGC GGCGCGGCAA TATTTTGCAG CAGTTTTTGA TTGAGGCGGT
 951 GTTAATCTGC ATCATCGGAG GCTTGGTCGG CGTAGGTTTG TCCGCCGCCG
 1001 TCAGCCTCGT GTTCAATCAT TTTGTAACCG ATTTCCCGAT GGACATTTCC
 1051 GCGGCATCCG TTATCGGGGC GGTCCGCTGT TCGACCGGAA TCGGCATCGC
 55 1101 GTTCGGCTTT ATGCCTGCCA ATAAGGCAGC CAACTCAAT CCGATAGATG
 1151 CATTGGCGCA GGATTGA

This encodes a protein having amino acid sequence <SEQ ID 538>:

1 MSVQAVLAHK MRSLLTMLGI IIGIASVSV VALNGSQKK ILEDISSMGT
 51 NTISIFPGRG FGDRRSKIK TLTIDDAKII AKQSYVASAT PMTSSGGTLT
 60 101 YRNTDLTASL YGVGEQYFDV RGLKLETGRL FDENDVKEDA QVVVIDQNVK
 151 DKLFADSDPL GKTILFRKRP LTVIGVMKKD ENAFGNSDVL MLWSPYTTVM
 201 HQITGESHTN SITVKIKDNA NTRVAEKGLE ELLKARHGTE DFFMNSDSI
 251 RQMVESTTGT MKLLISSIAL ISLVVGGIGV MNIMLVSVTE RTKEIGIRMA
 301 IGARRGNILQ QFLIEAVLIC IIGGLVGVGL SAAVSLVFNH FVTDFPMDIS

351 AASVIGAVAC STGIGIAFGF MPANKAAKLN PIDALAQD*

ORF134ng and ORF134-1 show 97.9% identity in 388 aa overlap:

```

5   orf134ng      MSVQAVLAHKMRSLLTMLGIIIGIASVSVVALNGSQKKILEDISSMGNTNTISIFPGRG
    orf134-1      MSVQAVLAHKMRSLLTMLGIIIGIASVSVVALNGSQKKILEDISSIGTNTISIFPGRG

10  orf134ng      FGDRRSKGIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV
    orf134-1      FGDRRSKGRIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV

15  orf134ng      RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTLFRKRPLTVIGVMKKD
    orf134-1      RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFADSDPLGKTLFRKRPLTVIGVMKKD

20  orf134ng      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGT
    orf134-1      ENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTQVAEKGLTDLKARHGT

25  orf134ng      DFFMNNSDSIRQMVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA
    orf134-1      DFFMNNSDSIRQIVESTTGTMKLLISSIALISLVVGGIGVMNIMLVSVTERTKEIGIRMA

    orf134ng      IGARRGNILQQFLIEAVLICIIIGLVGVGLSAAVSLVFNHFVTDFFPMDISAASVIGAVAC
    orf134-1      IGARRGNILQQFLIEAVLICVIGLVGVGLSAAVSLVFNHFVTDFFPMDISAMSVIGAVAC

    orf134ng      STGIGIAFGFMPANKAAKLNPIDALAQDX
    orf134-1      STGIGIAFGFMPANKAAKLNPIDALAQDX

```

30 ORF134ng also shows homology to an *E.coli* ABC transporter:

sp|P75831|YBJZ ECOLI HYPOTHETICAL ABC TRANSPORTER ATP-BINDING PROTEIN YBJZ >gi5
(AE000189) o648; similar to YBBA_HAEIN SW: P45247 [Escherichia coli] Length = 648

Score = 297 bits (753), Expect = 6e-80
Identities = 162/389 (41%), Positives = 230/389 (58%), Gaps = 1/389 (0%)

```

35  Query: 1  MSVQAVLAHKMRSLLTMLXXXXXXXXXXXXXXXXXGNGSQKKILEDISSMGNTNTISIFPGRG 60
    M+ +A+ A+KMR+LLTML +G+ +++ +L DI S+GTNTI ++PG+
    Sbjct: 260 MAWRALAANKMRTLLTMLGIIIGIASVSVVVGDAAKQMVLAIRSIGTNTIDVYPGKD 319

40  Query: 61  FGDRRSKGIKTLTIDDAKIIAKQSYVASATPMTSSGGTLTYRNTDLTASLYGVGEQYFDV 120
    FGD + L DD I KQ +VASATP S L Y N D+ AS GV YF+V
    Sbjct: 320 FGDDDPQYQQALKYDDLIAIQKPWVASATPAVSQNLRLRYNNVDVAASANGVSGDYFNV 379

45  Query: 121 RGLKLETGRLFDENDVKEDAQVVVIDQNVKDKLFAD-SDPLGKTLFRKRPLTVIGVMKK 179
    G+ G F++ + AQVVV+D N + +LF +D +G+ IL P VIGV ++
    Sbjct: 380 YGMTFSEGNTFNQEQLNGRAQVVVLDNTRRQLFPHKADVGEVILVGNMPPARVIGVAEE 439

50  Query: 180 DENAFGNSDVLMLWSPYTTVMHQITGESHTNSITVKIKDNANTRVAEKGLAELLKARHGT 239
    ++ FG+S VL +W PY+T+ ++ G+S NSITV++K+ ++ AE+ L LL RHG
    Sbjct: 440 KQSMFGSSKVLRLVWLPYSTMSGRVMGQSWLNSITVRVKEGFDSEAEQQLTRLLSLRHGK 499

55  Query: 240 EDFFMNNSDSIRQMVESTTGTMKXXXXXXXXXXXXXVGGIGVMNIMLVSVTERTKEIGIRM 299
    +DFF N D + + VE TT T++ VGGIGVMNIMLVSVTERT+EIGIRM
    Sbjct: 500 KDFFTWNMDGVLKTEKTRTLQLFLTLVAVISLVVGGIGVMNIMLVSVTERTREIGIRM 559

60  Query: 300 AIGARRGNILQQFLIEXXXXXXXXXXXXXXXXXXXXXFNHFVTDFFPMDISAASVIGAVA 359
    A+GAR ++LQQFLIE F+ + + S +++ A
    Sbjct: 560 AVGARASDVLQQFLIEAVLVCLVGGALGITLSLLIAFTLQLFLPGWEIGFSPLALLLAF 619

    Query: 360 CSTGIGIAFGFMPANKAAKLNPIDALAQD 388
    CST GI FG++PA AA+L+P+DALA++
    Sbjct: 620 CSTVTGILFGWLPARNAARLDPVDALARE 648

```

Based on this analysis, including the presence of the leader peptide and transmembrane regions in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 65

- 5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 539>:

```

      1 ..GGGACGGGAG CGATGCTGCT GCTGTTTTAC GCGGTAACGA T.CTGCCTTT
    51 GGCCACTGGC GTTACCCTGA GTTACACCTC GTCGATTTTT TTGGCGGTAT
   101 TTTCCTTCCT GATTTTGAAA GAACGGATTT CCGTTTACAC GCAGGCGGTG
   151 CTGCTCCTTG GTTTTGCCGG CGTGGTATTG CTGCTTAATC CCTCGTTCGG
   201 CAGCGGTCAG GAAACGGCGG CACTCGCCGG GCTGGCGGGC GGC CGCATGT
   251 CCGGCTGGGC GTATTTGAAA GTGCGCGAAC TGTCTTTGGC GGGCGAACCC
   301 GGCTGGCGCG TCGTGTTTTA CCTTTCCGTG ACAGGTGTGG CGATGTCGTC
   351 GGTTTGGGCG ACGCTGACCG GCTGGCACAC CCTGTCCTTT CCATCGGCAG
   401 TTTATCTGTC GTGCATCGGC GTGTCGCGC TGATTGCCCA ACTGTCGATG
   451 ACGCGCGCCT ACAAAAGTCGG CGACAAATTC ACGGTTGCCT CGCTTTCCTA
   501 TATGACCGTC GTTTTTCCG CTCTGTCTGC CGCATTTTTT CTGGGCGAAG
   551 AGCTTTTCTG GCAGGAAATA CTCGGTATGT GCATCATCAT CCTCAGCGGT
   601 ATTTTGA
  
```

This corresponds to the amino acid sequence <SEQ ID 540; ORF135>:

```

  20      1 ..GTGAMLLFY AVTILPLATG VTLSTSSIF LAVFSFLILK ERISVYTQAV
      51 LLGFAGVVL LLNPSFRSGQ ETAALAGLAG GAMSGWAYLK VRELSLAGEP
   101 GWRVVFYLSV TGVAMSSVWA TLTGWHTLSF PSAVYLSCIG VSALIAQLSM
   151 TRAYKVGDKF TVASLSYMTV VFSALSAAFF LGEELEWQEI LGMCIISAV
   201 F*
  
```

- 25 Further work revealed the complete nucleotide sequence <SEQ ID 541>:

```

      1 ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGCG
      51 GGCGGCCTGC TTTACCATTG TGAACGTATT GATTAAAGAG GCATCGGCAG
   101 AATTGCCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTTCA
   151 ACCGTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA mCTTCCGCAC
   201 GCCCCATTGG AAAAACCACT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
   251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGGC CACTGGCGTT
   301 ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCTGAT
   351 TTTGAAAGAA CGGATTTCGG TTTACACGCA GGCGGTGCTG CTCCTTGGTT
   401 TTGCCGGCGT GGTATTGCTG CTTAATCCCT CGTTCCGCAG CGGTCAAGAA
   451 ACGGCGGCAC TCGCCGGGCT GGCGGGCGGC GCGATGTCCG GCTGGGCGTA
   501 TTTGAAAGTG CGCGAACTGT CTTTGGCGGG CGAACCCGGC TGGCGCGTCG
   551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCGTCGGT TTGGGCGACG
   601 CTGACCGGCT GGCACACCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCGTG
   651 CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA
   701 AAGTCGCGCA CAAATTCACG GTTGCCTCGC TTTCTATAT GACCGTCGTT
   751 TTTTCCGCTC TGTCTGCCGC ATTTTTTCTG GGCGAAGAGC TTTTCTGGCA
   801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
   851 TCCGCCCCAC TGCCTTCAAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
   901 TAA
  
```

- 45 This corresponds to the amino acid sequence <SEQ ID 542; ORF135-1>:

```

      1 MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRLFS
      51 TVALGAAAVL RRDxFRTPHW KNHLNRSVMG TGAMLLLFYA VTHLPLATGV
   101 TLSYTSSIFL AVFSFLILKE RISVYTOAVL LLGFAGVLL LNPSFRSGQE
   151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVMSSVWAT
   201 LTGWHTLSFP SAVYLSCIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
   251 FSALSAAFFL GEELFWQEI GMCTIILSGI LSSIRPTAFK QRLQSLFRQR
   301 *
  
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF135 shows 99.0% identity over a 197aa overlap with an ORF (ORF135a) from strain A of *N.*

meningitidis:

```

5      orf135.pep      10      20      30
                        GTGAMLLLFYAVTILPLATGVTLSYSSIF
                        |||
orf135a      STVALGAAAVLRRDFTFTPHWKNHLNRSMTGAMLLLFYAVTHLPLATGVTLSYSSIF
                50      60      70      80      90      100

10     orf135.pep      40      50      60      70      80      90
                        LAVFSFLILKERISVYTQAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK
                        |||
orf135a      LAVFSFLILKERISVYTQAVLLLGAGVLLNPSFRSGQETAALAGLAGGAMSGWAYLK
                110     120     130     140     150     160

15     orf135.pep      100     110     120     130     140     150
                        VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFSPSAVYLSICIGVSALIAQLSM
                        |||
orf135a      VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFSPSAVYLSICIGVSALIAQLSM
                170     180     190     200     210     220

20     orf135.pep      160     170     180     190     200
                        TRAYKVGDKFTVASLSYMTVVFSALSAFFLGEELFWQEIILGMCIIISAVFX
                        |||
orf135a      TRAYKVGDKFTVASLSYMTVVFSALSAFFLAEELFWQEIILGMCIIISGILSSIRPTAF
                230     240     250     260     270     280

orf135a      KQRLQSLFRQRX
                290     300

```

30 The complete length ORF135a nucleotide sequence <SEQ ID 543> is:

```

1  ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGCC
51  GGCGGCCTGC TTTACCATT TGAACGTATT GATTAAAGAG GCATCGGCAG
101 AATTGCCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTCA
151 ACCGTTGCGC TCGGGGCTGC CGCCGTATTG CGTCGGGACA CCTTCCGCAC
201 GCCCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
251 TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGGC CACCGGCGTT
301 ACCCTGAGTT ACACCTCGTC GATTTTTTTG GCGGTATTTT CCTTCCTGAT
351 TTTGAAAGAA CGGATTTCGG TTTACACGCA GGCGGTGCTG CTCCTTGGTT
401 TTGCCGGCGT GGTATTGCTG CTTAATCCCT CGTTCGCGAG CCGTCAGGAA
451 ACGGCGGCAC TCGCCGGGCT GGCGGGCGGC GCGATGTCGG GCTGGGCGTA
501 TTTGAAAGTG CGCGAAGTGT CTTTGGCGGG CGAACCAGGC TGGCGCGTCG
551 TGTTTTACCT TTCCGTGACA GGTGTGGCGA TGTCATCGGT TTGGGCGACG
601 CTGACCGGCT GGCACACCCT GTCCTTTCCA TCGGCAGTTT ATCTGTCGTG
651 CATCGGCGTG TCCGCGCTGA TTGCCCAACT GTCGATGACG CGCGCCTACA
701 AAGTCGGCGA CAAATTCACG GTTGCCTCGC TTCCTATAT GACCGTCGTT
751 TTTTCCGCTC TGTCTGCCGC ATTTTCTCTG GCCGAAGAGC TTTTCTGGCA
801 GGAAATACTC GGTATGTGCA TCATCATCCT CAGCGGTATT TTGAGCAGCA
851 TCCGCCCCAC TGCCTTCAAA CAGCGGCTGC AATCCCTGTT CCGCCAAAGA
901 TAA

```

50 This encodes a protein having amino acid sequence <SEQ ID 544>:

```

1  MDTAKKDILG SGWMLVAAAC FTIMNVLIKE ASAKFALGSG ELVFWRMFLS
51  TVALGAAAVL RRDFTFTPHW KNHLNRSMTG TGMALLFYA VTHLPLATGV
101 TLSYSSIFL AVFSEFLIKE RISVYTQAVL LLGFAGVLL LNPFSRSGQE
151 TAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSVT GVAMSSVWAT
201 LTGWHTLSFP SAVYLSICIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
251 FSALSAFFL AEELFWQEIL GMCIIILSGI LSSIRPTAFK QRLQSLFRQR
301 *

```

ORF135a and ORF135-1 show 99.3% identity in 300 aa overlap:

```

60  orf135a.pep      MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL
                        |||
orf135-1      MDTAKKDILGSGWMLVAAACFTIMNVLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL

```

5	orf135a.pep	RRDTFRTPHWKHNLRSMVGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE
	orf135-1	RRDXFRTPHWKHNLRSMVGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE
10	orf135a.pep	RISVYTQAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
	orf135-1	RISVYTQAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
15	orf135a.pep	WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
	orf135-1	WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
20	orf135a.pep	VASLSYMTVVFSAALAEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR
	orf135-1	VASLSYMTVVFSAALAEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR

Homology with a predicted ORF from *N.gonorrhoeae*

ORF135 shows 97% identity over a 201aa overlap with a predicted ORF (ORF135ng) from

20	orf135.pep	GTGAMLLLFYAVTXLPLATGVTLSTSSIF 30
	orf135ng	STVTLGAAVLRDRTFRTPHWKHNLRSMVGTGAMLLLFYAVTHLPLTTGVTLSTSSIF 335
25	orf135.pep	LAVFSFLILKERISVYTQAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLK 90
	orf135ng	LAVFSFLILKERISVYTQAVLLLGAGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLK 395
30	orf135.pep	VRELSLAGEPGWRVVFYLSVTGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM 150
	orf135ng	VRELSLAGEPGWRVVFYLSATGVAMSSVWATLTGWHTLSFPSAVYLSGIGVSALIAQLSM 455
35	orf135.pep	TRAYKVGDKFTVASLSYMTVVFSAALAEELFWQEILGMCIIISAVF 201
	orf135ng	TRAYKVGDKFTVASLSYMTVVFSAALAEELFWQEILGMCIIISAAF 506

An ORF135ng nucleotide sequence <SEQ ID 545> was predicted to encode a protein having amino acid sequence <SEQ ID 546>:

40	1	MPSEKAFRRH LRTASFOGLH LHHFHQKVGK CGIIGFGIHI FPTLLPAAQG
	51	ILDIOQLGLFR IDFAALAVYR RTQVDFIHTV IDGIASDQAF SEVVQILRRL
45	101	NLGHFTDTHL IAQARRFIAD FGNIRPMRRG EAKTFCRCFR FDGIDGIHGD
	151	FRQCGHINRL APGKDCRNGK RDKVFFHTRH YNQVCLEKTN CSARKIKFRH
50	201	QKQAKTHSTS LAARFTIRPS LSQRPFMDTA KKDILGSGWM LVAAACFTVM
	251	NVLIKEASAK FALGSGELVF WRMLFSTVTL GAAAVLRDRT FRTPHWKHNL
55	301	NRSMVGTGAM LLLFYAVTHL PLTTGVTLST TSSIFLAVFS FLILKERISV
	351	YTOAVLLLG AGVVLLNPS FRSGQEPAL AGLAGGAMSG WAYLKVRELS
60	401	LAGEPGWRVV FYLSATGVAM SSVWATLTGW HTLSFPSAVY LSGIGVSALI
	451	AQLSMTRAYK VGDKFTVASL SYMTVVFSA SAAFFLGEEL FWQEILGMC
	501	IISAAF*

Further work revealed the following gonococcal sequence <SEQ ID 547>:

50	1	ATGGATACCG CAAAAAAGA CATTTTAGGA TCGGGCTGGA TGCTGGTGGC
	51	GGCGGCCTGC TTCACCGTTA TGAACGTATT GATTAAAGAG GCATCGGCAA
55	101	AATTTGCCCT CGGCAGCGGC GAATTGGTCT TTTGGCGCAT GCTGTTTCA
	151	ACCGTTACGC TCGGTGCTGC CGCGGTATTG CGGCAGCACA CCTTCCGCAC
60	201	GCCCCATTGG AAAAACCCT TAAACCGCAG TATGGTCGGG ACGGGGGCGA
	251	TGCTGCTGCT GTTTTACGCG GTAACGCATC TGCCTTTGAC AACCAGCGTT
65	301	ACCCTGAGTT ACACCTCGTC GATTTTttg GCGGTATTTT CCTTCTGAT
	351	TTTGAAAGAA CGGATTTCG TTTACACGCA GCGGTGCTG CTCCTTGTT
70	401	TTGCCGGCGT GGTATTGCTG CTTAATCCCT CGTCCGCAG CGGTGAGAA
	451	CCGGCGGCAC TCGCCGGGCT GCGGGGCGGC GCGATGTCG GCTGGGCGTA
75	501	TTTGAAAGTG CGCGAATCTG CTTTGGCGGG CGAACCCGGC TGCGCGGTCG
	551	TGTTTACCT TTCCGCAACC GCGGTGGCGA TGTCGTCggt ttgggagacg
	601	Ctgaccggct ggCACaccct GTCCTTcca tcggcagttt ATCtgtCGG

5 651 CATCGGCGTG tccgcgCtgA TTGCCCAaCT GtcgatgAcg CGCGcctaca
 701 aaGTTCGGCGA CAAATTCACG GTTGCTCGC tttcctaTat gaccgtcGTC
 751 TTTTCGCCCC TGTCTGCCG ATTTTCTCTg ggcgaagagc tttTctggCA
 801 GGAAATACTC GGTATGTGCA TCATTatccT CAGCGGCATT TTGAGCAGCA
 851 TCCGCCCCAT TGCCTTCAA CAGCGGCTGC AAGCCCTCTT CCGCCAAAGA
 901 TAA

This corresponds to the amino acid sequence <SEQ ID 548; ORF135ng-1>:

10 1 MDTAKKDILG SGWMLVAAAC FTMNVNLIKE ASAKFALGSG ELVFWRMFLS
 51 TVTLGAAAVL RRDFTFRPHW KNHLNRSVMG TGAMLLLFYA VTHLPLTTGV
 101 TLSYTSSIFL AVFSFLILKE RISVYTQAVL LLGFAGVLL LNPSFRSGQE
 151 PAALAGLAGG AMSGWAYLKV RELSLAGEPG WRVVFYLSAT GVAMSSVWAT
 201 LTGWHTLSFP SAVYLSGIGV SALIAQLSMT RAYKVGDKFT VASLSYMTVV
 251 FSALSAFFL GEELFWQEIL GMCIIILSGI LSSIRPIAFK QRLQALFRQR
 301 *

15 ORF135ng-1 and ORF135-1 show 97.0% identity in 300 aa overlap:

20 orf135ng-1.pep MDTAKKDILGSGWMLVAAACFTVMNVNLIKEASAKFALGSGELVFWRMFLSTVTLGAAAVL
 orf135-1 MDTAKKDILGSGWMLVAAACFTIMNVNLIKEASAKFALGSGELVFWRMFLSTVALGAAAVL
 25 orf135ng-1.pep RRDFTFRPHWKNHLNRSVMGTGAMLLLFYAVTHLPLTTGVTLSTSSIFLAVFSFLILKE
 orf135-1 RRDXTFRPHWKNHLNRSVMGTGAMLLLFYAVTHLPLATGVTLSTSSIFLAVFSFLILKE
 30 orf135ng-1.pep RISVYTQAVLLLGFGVLLLNPSFRSGQEPALAGLAGGAMSGWAYLKVRELSLAGEPG
 orf135-1 RISVYTQAVLLLGFGVLLLNPSFRSGQETAALAGLAGGAMSGWAYLKVRELSLAGEPG
 35 orf135ng-1.pep WRVVFYLSATGVAMSSVWATLTGWHTLSFPSSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
 orf135-1 WRVVFYLSVTGVAMSSVWATLTGWHTLSFPSSAVYLSGIGVSALIAQLSMTRAYKVGDKFT
 40 orf135ng-1.pep VASLSYMTVVFSALSAFFLGEELFWQEILGMCIIILSGILSSIRPIAFKQRLQALFRQR
 orf135-1 VASLSYMTVVFSALSAFFLGEELFWQEILGMCIIILSGILSSIRPTAFKQRLQSLFRQR

35 Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 66

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 549>:

40 1 ATGAAGCGGC GTATAGCCGT CTTGCTCCTG TTCCCGCAGA TAATCCGAGT
 51 TTTGGGACAA CTGTTGCCGA AAATCGTCAA TACAGTTCCG GCACATCGGA
 101 TGCTCTTCCA GATTTCGGG ATGTTCTTT TCTTCATACA CCAGCAATAT
 151 CTGCCCCGGA TCGCCGAAAT CGATTCCCA TCGGCATCG TGTTCGGTGC
 201 GCTCCTCTTC CGTCATCTGC CCGCGCATTG CCTGTATGGT AAAGCCGCCG
 45 251 TAGGGGATGC CgTTGCACAC GAACATCCAG TCGCTGATGT CGTCAACCGG
 301 AACGCAAACG cTTTCGCCTT GTTCGACATT GGTGAGTTTC CCsGGTTCAT
 351 TGTTACGAC ACCGTAAATA TAAAGACCGT CAAAATAAAT ATCGTCGATC
 401 CACATATGTT CGCAAATTTC GCCGTCTTCG CCGTCTTGA AAAAAGGGAC
 451 TTTGACCATG GCAAAATCCA AGCGGAAAT AATCGGCGCG CGTTCCCAAA
 50 501 AAAGcTCGCG CCAAAATAT TTGAATGTTT TACGGGCGCG TTCGTGGCA
 551 CGGTTTACCG GTTCGCTGTC CTGTTCTACA TAATAAATGA CGGAATCGCC
 601 CATCATATCT GCTCCTCAAC GTGTACGGTA TCTGTTGCA CCTTACTGCG
 651 GCTTTCTgCc KTCGGCATCC GATTGCGATT TGAAAAGTTC mmrwyATTCC
 701 GAATAG

55 This corresponds to the amino acid sequence <SEQ ID 550; ORF136>:

1 MKRRIAVEVL FPQIIRVLGQ LLPKIVNTVP AHRMLFQIFG MFFFFIHQQY
 51 LPGIAEIDSP CGIVFGALLF RELPAHCLYG KAAVGDAVAH EHPVADVVR

-320-

101 NANAFALFDI GQFAXFIVQH TVNIKTVKIN IVDPHMFANF AVFAVLEKRD
 151 FDHGKIQGGN NAAAFPKKLA PKIFECFTGA FVGTVYRFVC LFYIINDGIA
 201 HHSAPQVRVY LFAPYCGFLP SASDSLKSS XXSE*

Further work revealed the complete nucleotide sequence <SEQ ID 551>:

5 1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGTTCCCGC AGATAATCCG
 51 AGTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC
 101 101 GGATGCTCTT CCAGATTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA
 151 TATCTGCCCG GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG
 201 201 TCGCTCCTC TTCCGTCATC TGCCCGCGCA TTGCTGTAT GGTAAAGCCG
 10 251 CCGTAGGGGA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCGTCAAC
 301 CGGAACGCAA ACGCTTTTCGC CTGTTCGAC ATTGGTCAGT TCGCCGGGTT
 351 CATTGTTCAG CACACCGTAA ATATAAAGAC CGTCAAAATA AATATCGTCG
 401 ATCCACATAT GTTCGCAAT TTCCGCGTCT TCGCCGCTCTT GGAAAAAAGG
 451 451 GACTTTGACC ATGGCAAAAT CCAAGGCGGA AATAATGCGG CGCGGTTCCC
 15 501 AAAAAAGCTC GCGCCAAAAA TATTGAATG TTTTACGGGC GCGTTCGTCG
 551 GCACGGTTTA CCGGTTTCGTC TGCTGTCTT ACATAATAA TGACGGAATC
 601 GCCCATCATT CTGCTCCTCA ACGTGTACGG TATCTGTTG CACCTTACTG
 651 CGGCTTTCG CCTTCGGCAT CCGATTCGGA TTTGAAAAGT TCCAAATATT
 701 701 CGGAATAG

20 This corresponds to the amino acid sequence <SEQ ID 552; ORF136-1>:

 1 MMKRRIAVFV LFPQIIRVLG QLLPKIVNTV PAHRMLFQIF GMFFFFIHQQ
 51 YLPGIAEIDS PCGIVFGALL FRHLPAHCLY GKAAGDAVA HEHPVADVVN
 101 101 RNANAFALFD IGQFAGFIVQ HTVNIKTVKI NIVDPHMFAN FAVFAVLEKR
 151 151 DFDHGKIQGG NNAAFPKKL APKIFECFTG AFVGTVYRFV CLFYIINDGI
 201 201 AHHSAPQVRV YLFAPYCGFL PSASDSLKS SKYSE*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF136 shows 71.7% identity over a 237aa overlap with an ORF (ORF136a) from strain A of *N. meningitidis*:

30	orfl36.pep	10	20	30	40	50	59
		MKRRIAVFVLFPQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGIAEIDS					
	orfl36a	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLPGIAEIDS					
35	orfl36.pep	60	70	80	90	100	119
		PCGIVFGALLFRHLPAHCLYGYKAAVGDAVAHEHPVADVVRNANAFALFDIGQFAXFIVQ					
40	orfl36a	PCGIVFGTLLFRHXSTHCLYGYKAAVGNAVAHEHPVADVVRNANAFALFDIGQFAGFIVQ					
45	orfl36a	130	140	150	160	170	180
		HAINVKTVKINIVDPHMFANFAXFAVLEKRALTMASKSKXXMRRRSQKSSRQKYNVLR					
50	orfl36a	180	190	200	210	220	230
		AFVGTVYRFVCLFYIINDGIAHH---SAPQVRVYLFAPYCGFLPSASDSLKSSXXSEX					
		R---SPARFTGLSACSTXXMTESPIISAPQVRVYLFAPYCGFLPSASDSLKSSKYSEX					

The complete length ORF136a nucleotide sequence <SEQ ID 553> is:

55 1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG
 51 GATTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC
 101 101 GGATGCTCTT CCAGATNTTC GGGATGTTCT TTTTCTTCAT ACACCAGCAA
 151 TACCTGCCCG GGATCGCCGA AATCGATTCC CCATGCGGCA TCGTGTTCGG
 201 201 TACGCTCCTC TTCCGTCATC NGTCCACGCA TTGCTGTAT GGTAAAGCCG
 251 251 CCGTAGGGAA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCGTCAAC

-321-

5
301 CGGAACGCAA ACGCTTTCGC CTTGTTTCGAC ATTGGTCAGT TCGCCGGGTT
351 CATTGTTTCAG CACGCCATAA ATGTAAAGAC CGTCAAAATA AATATCGTCG
401 ATCCACATAT GTTCGCAAAAT TTCGCCNTCT TCGCCGTCTT GGAAAAAAGG
451 GCTTTGACCA TGGCAAAATC TAAGGNGNNA NNGATGCGGC GCGGTTCCCA
501 AAAAAGCTCG CGCCAAAAAT ATTTGAATGT TTTGCGGGCG CGTTCGCCGG
551 CACGGTTTAC CGGTTTGTCT GCCTGTCTTA CATAATAAT GACGGAATCG
601 CCCATCATAT CTGCTCCTCA ACGTGTACGG TATCTGTTTG CACCTTACTG
651 CGGCTTTCTG CCTTCGGCAT CCGATTCGGA TTTGAAAAGT TCCAAATATT
701 CGGAATAG

10 This encodes a protein having amino acid sequence <SEQ ID 554>:

1 MMKRRIAVFV LLMQKIRILG QLLPKIVNTV PAHRMLFQXF GMFFFFIHQQ
51 YLPGIAEIDS PCGIVFGTLL FRHXSTHCLY GKAAVGNAVA HEHPVADVNV
101 RNANAFALFD IGQFAGFIVQ HAINVKT VKI NIVDPHMEAN FAXFAVLEKR
151 ALTMASKXX XMRRRSQKSS RQKYLNLRA RSPARFTGLS ACST**MTES
201 PIISAPQVR YLFAPYCGFL PSASDSLKS SKYSE*

ORF136a and ORF136-1 show 73.1% identity in 238 aa overlap:

		10	20	30	40	50	60
orfl36a.pep		MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQXFGMFFFFIHQQYLP					
20	orfl36-1	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLP					
		10	20	30	40	50	60
		70	80	90	100	110	120
25	orfl36a.pep	PCGIVFGTLLFRHXSTHCLYGKAAVGNAVAHEHPVADVNVNRNANAFALFDIGQFAGFIVQ					
	orfl36-1	PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVNVNRNANAFALFDIGQFAGFIVQ					
		70	80	90	100	110	120
		130	140	150	160	170	180
30	orfl36a.pep	HAINVKT VKI NIVDPHMEAN FAXFAVLEKRALTMASKXXXMRRRSQKSSRQKYLNLRA					
	orfl36-1	HTVNIKT VKI NIVDPHMEAN FAXFAVLEKRDFDHGKIQGGNNAAFPKKLAPKIFECFTG					
		130	140	150	160	170	180
35		190	200	210	220	230	
	orfl36a.pep	R---SPARFTGLSACSTXXMTESPIISAPQVRVYLFAPYCGFLPSASDSLKS					
	orfl36-1	AFVGTVYRFVCLFYIINDGIAHH---SAPQVRVYLFAPYCGFLPSASDSLKS					
		190	200	210	220	230	

Homology with a predicted ORF from *N.gonorrhoeae*

ORF136 shows 92.3% identity over a 234aa overlap with a predicted ORF (ORF136ng) from *N.gonorrhoeae*:

45	orfl36.pep	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLP	59
	orfl36ng	MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHQQYLP	60
	orfl36.pep	PCGIVFGALLFRHLPAHCLYGKAAVGDAVAHEHPVADVNVNRNANAFALFDIGQFAXFIVQ	119
50	orfl36ng	PGGIVFGTLLFRHLSAHCLYGKAAVGDAVAHEHPVADVNVNRNANAFALFDIGQSAGFIVQ	120
	orfl36.pep	HTVNIKT VKI NIVDPHMEAN FAXFAVLEKRDFDHGKIQGGNNAAFPKKLAPKIFECFTG	179
	orfl36ng	HTVNIKT VKI NIVDPHMEAN FAXFAVLEKRDFDHGKIQGGNNAAFPKKLAPKVFEFTG	180
55	orfl36.pep	AFVGTVYRFVCLFYIINDGIAHHSAPQVRVYLFAPYCGFLPSASDSLKS	234
	orfl36ng	AFAGTVYRFVCLFYIINDGIAHTAPQVRVYLFAPYRGFLPPASDSLKS	235

The complete length ORF136ng nucleotide sequence <SEQ ID 555> is:

60
1 ATGATGAAGC GCGGTATAGC CGTCTTCGTC CTGCTCATGC AGAAAATCCG
51 GATTTTGGGA CAACTGTTGC CGAAAATCGT CAATACAGTT CCGGCACATC

-322-

101 GGATGCTCTT CCAAATTTTC GGGATGTTCT TTTTCTTCAT ACACCGGCAA
 151 TACCTGCCCG GGATCGCCGA AATCGATTCC CCAGGCGGTA TCGTGTTCCG
 201 TACGCTCCTC TTCCGTCATC TGTCGCGCA TTGCTGTAC GGTAAGCCG
 251 CCGTAGGGGA TGCCGTTGCA CACGAACATC CAGTCGCTGA TGTCGCCAAC
 301 CGGAACGCAA ACGCTTTCGC CTTGTTCGAC ATTGTCAGT CCGCCGGGTT
 351 CATTGTTCAG CACACCGTAA ATATAAAGAC CGTCAAAATA AATATCGTCG
 401 ATCCACATAT GTTCGCAAAAT TTCGCCGTCT TCGCCGTCTT GGAAAAAAGG
 451 GACTTTGACC ATGGCAAAAT CCAAGGCGGA AATAATGCGG CGGCGTTCCC
 501 AAAAAAGCTC GCGCCAAAAG TATTTGAATG TTTTACGGGC GCGTTCGCCG
 551 GCACGGTTTA CCGGTTTCGTC TGCCGTGTTCT ACATAATAAA TGACGGGAATC
 601 GCCCATCATA CTGCTCCTCA ACGTGTACGG TATCTGTTTG CACCTTACCG
 651 CGGTTTCTA CCTCCGGCAT CCGATTCCGA TTTGAAAAGT TCCAATATT
 701 CGGAATAG

This encodes a protein having amino acid sequence <SEQ ID 556>:

15 1 MMKRRIAVFV LLMOKIRILG QLLPKIVNTV PAHRMLFQIF GMFFFFIHRQ
 51 YLPGIAEIDS PGGIVFGTLL FRHLSAHCLY GKAAVGDAVA HEHPVADVAN
 101 RNANAFALFD IGQSAGFIVQ HTVNIKT VKI NIVDPHMFAN FAVFAVLEKR
 151 DFDHGKIQQG NNAAPFKKL APKVFECFTG AFAGTVYRFV CLFYIINDGI
 201 AHHTAPQVRV YLFAPYRGFL PPASDSLKS SKYSE*

20 ORF136ng and ORF136-1 show 93.6% identity in 235 aa overlap:

orf136ng MMKRRIAVFVLLMQKIRILGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGIAEIDS
 orf136-1 MMKRRIAVFVLFQIIRVLGQLLPKIVNTVPAHRMLFQIFGMFFFFIHRQYLPGIAEIDS
 25 orf136ng PGGIVFGTLLFRHLSAHCLYGKAAVGDAVAHEHPVADVANRNANAFALFDIGQSAGFIVQ
 orf136-1 PCGIVFGALLFRHLPACLYGKAAVGDAVAHEHPVADVNRNANAFALFDIGQFAGFIVQ
 orf136ng HTVNIKT VKI NIVDPHMFANFAVFAVLEKRDFDHGKIQQGNNAAPFKKLAPKVFECFTG
 30 orf136-1 HTVNIKT VKI NIVDPHMFANFAVFAVLEKRDFDHGKIQQGNNAAPFKKLAPKIFECFTG
 orf136ng AFAGTVYRFVCLFYIINDGIAHHTAPQVRVYLFAPYRGFLPPASDSLKSSKYSEX
 35 orf136-1 AFVGTVYRFVCLFYIINDGIAHHSAPQVRVYLFAPYCGFLPSASDSLKSSKYSEX

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 67

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 557>:

1 ATGGAAAATA TGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC
 51 CGCCGCGCGG TTGCTTGCCG CC.TGCGGAC GGCGGGAAT AATGCTGTCC
 101 GCAAGCCGGT GCAAACCGCC AAACCGCGCG CAGTGGTCCG TTTGGCACTC
 151 GGTGGCGGCG CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
 45 201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACC TCCGCAGGTT
 251 CGATTGTCCG CAACCTTTTT GCATCGGGTA TGTCGCCCCG CCGCCTCGAA
 301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC
 351 CACCAATGGG TTTATCAAAG GCGCAAAGCT GCAAATAC ATCAACCGAA
 401 AACTCCGCGG CATGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCC...

50 This corresponds to the amino acid sequence <SEQ ID 558; ORF137>:

1 MENMVTFSKI RPLLAIAAAA LLAAXRTAGN NAVRKPVQTA KPAAVVGLAL
 51 GGGASKGFAH VGIIKVLKEN GIPVKVVTGT SAGSIVGNLF ASGMSPDRLE
 101 LEAEILGKTD LVDLTSTNG FIKGAKLQNY INRKLGMQI QQFPIKFAA..

Further work revealed the complete nucleotide sequence <SEQ ID 559>:

-323-

1 ATGGAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC
 51 CGCGGCCGCG TTGCTTGCCG CCTGCGGCAC GCGGGAAAT AATGCTGTCC
 101 GCAAGCCGGT GCAAACCGCC AAACCGCCG CAGTGGTCGG TTTGGCACTC
 151 GGTGGCGCGC CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
 201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT
 251 CGATTGTCGG CAGCCTTTTT GCATCGGGTA TGTCGCCCCG CCGCCTCGAA
 301 TTGGAAGCCG AAATTTTAGG CAAAACCGAT TTGGTCGATT TAACCTTGTC
 351 CACCAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
 401 AAGTCGGCGG CAGGCGGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT
 451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AGGGGAATGC
 501 CGGGCAGGCT GTGCGCGCTT CCGCGCCAT TCCCAATGTG TTCCAACCGG
 551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG
 601 CCCGTCAGTG CCGCCCGCGC GCAGGGGGCG AATTTCTGTA TTGCCGTCGA
 651 TATTTCCGCC CGTCCGGGCA AAAACATCAG CCAAGGTTTC TTCTCTTATC
 701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CTGCGTTGCA AAATGAGTTG
 751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGGATT TGGGTGCAGT
 801 CGGCGGATTC GATCAGAAAA AACCGCCAT CCGGTTGGGT GAGGAGGCAG
 851 CACGTGCCGC ATTGCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT
 901 TGA

20 This corresponds to the amino acid sequence <SEQ ID 560; ORF137-1>:

1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAVRKPVQTA KPAAVVGLAL
 51 GGGASKGFAH VGIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRL
 101 LEAEILGKTD LVDLTLSTSG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV
 151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQFVIGRHT YVDGGLSQPV
 201 PVSAARROGA NEVIAVDISA RPKNISOQF FSYLDQTLNV MSVSALQNEL
 251 GQADVVIKPO VLDLGAUGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY
 301 *

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

30 ORF137 shows 93.3% identity over a 149aa overlap with an ORF (ORF137a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
orf137.pep	MENMVTFSKIRPLLAIAAAA	LLAAXRTAGN	NAVRKPVQTA	KPAAVVGLAL	GGGASKGFAH		
orf137a	MENMVTFSKIRPLLAIAAAA	LLAACGTAGN	NAARKPVQTA	KPAAVVGLAL	GGGASKGFAH		
		10	20	30	40	50	60
orf137.pep	VGIKVLKENGIPVKVVTG	SAGSIVGNLF	ASGMSPDRL	LEAEILGKTD	LVDLTLSTNG		
orf137a	VGIKVLKENGIPVKVVTG	SAGSIVGSLF	ASGMSPDRL	LEAEILGKTD	LVDLTLSTSG		
		70	80	90	100	110	120
orf137.pep	FIKGAQLQNYINRKL	RGMQIQFPIKFAA					
orf137a	FIKGEKLQNYINRKVG	RRIQFPIKFAAV	ATDFETGKAV	AFNQGNAGQ	AVRASAAIPNV		
		130	140	150	160	170	180

The complete length ORF137a nucleotide sequence <SEQ ID 561> is:

50 1 ATGGAATA TGGTAACGTT TTCAAAAATC AGACCGCTTT TGGCAATCGC
 51 CGCGGCCGCG TTGCTTGCCG CCTGCGGCAC GCGGGAAAT AATGCTGTCC
 101 GCAAGCCGGT GCAAACCGCC AAACCGCCG CAGTGGTCGG TTTGGCACTC
 151 GGTGGCGCGC CATCTAAAGG ATTTGCCCAT GTAGGTATTA TTAAGGTTTT
 201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT
 55 251 CGATAGTCGG CAGCCTTTTT GCATCGGGTA TGTCGCCCCG CCGCCTCGAA
 301 TTGGAAGCCG AAATTTTAGG TAAAACCGAT TTGGTCGATT TAACCTTGTC
 351 CACCAGTGGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
 401 AAGTCGGCGG CAGGCGGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT
 451 GCTACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AAGGGGAATGC
 501 CGGGCAGGCT GTGCGCGCTT CCGCGCCAT TCCCAATGTG TTCCAACCGG
 60 551 TTATCATCGG CAGGCATACA TATGTTGACG GCGGTCTGTC GCAGCCCGTG

-324-

5
601 CCGGTCAGTG CCGCCCGGCG GCANGNNNG NATNTCGTGA TTGCCGTCGA
651 TATTTCCGCC CGTCCGAGCA AAAACATCAG CCAAGGCTTC TTCTCTTATC
701 TCGATCAGAC GCTGAACGTA ATGAGCGTTT CCGCGTTGCA AAATGAGTTG
751 GGGCAGGCGG ATGTGGTTAT CAAACCGCAG GTTTTGATT TGGGTGCAGT
801 CGGCGGATTC GATCAGAAAA AACGCGCCAT CCGGTTGGGT GAGGAGGCAG
851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT
901 TGA

This encodes a protein having amino acid sequence <SEQ ID 562>:

10
1 MENMVTFSKI RPLLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVGLAL
51 GGGASKGFAH VGIKVLKEN GIPVKVVTGT SAGSIVGSLF ASGMSPDRL
101 LEAEILGKTD LVDLTSTSG FIKGEKLQNY INRKVGGRRI QQFPIKFAAV
151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHT YVDGGLSQPV
201 PVSAARRXXX XXVIAVDISA RPSKNISQGF FSYLDQTLNV MSVSALQNEL
251 QQADVVIKPO VLDLGAUGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY
15 301 *

ORF137a and ORF137-1 show 97.3% identity in 300 aa overlap:

20
orf137a.pep MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVGLALGGGASKGFAH
orf137-1 MENMVTFSKIRPLLAIAAAAALLAACGTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH
orf137a.pep VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLLEAEILGKTDLVDLTSTSG
orf137-1 VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLLEAEILGKTDLVDLTSTSG
25
orf137a.pep FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
orf137-1 FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
30
orf137a.pep FQPVIIGRHTYVDGGLSQPVVSAARRXXXXXVIAVDISARPSKNISQGFYSYLDQTLNV
orf137-1 FQPVIIGRHTYVDGGLSQPVVSAARRQGANFVIAVDISARPGKNISQGFYSYLDQTLNV
35
orf137a.pep MSVSALQNELQADVVIKPOVLDLGAUGGFQDQKKRAIRLGEEAARAALPEIKRKLAAARY
orf137-1 MSVSALQNELQADVVIKPOVLDLGAUGGFQDQKKRAIRLGEEAARAALPEIKRKLAAARY

Homology with a predicted ORF from *N.gonorrhoeae*

ORF137 shows 89.9% identity over a 149aa overlap with a predicted ORF (ORF137ng) from *N.gonorrhoeae*:

40
orf137.pep MENMVTFSKIRPLLAIAAAAALLAAXRTAGNNAVRKPVQTAKPAAVVGLALGGGASKGFAH 60
orf137ng MENMVTFSKIRSFLLAIAAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGGASKGFAH 60
45
orf137.pep VGIKVLKENGIPVKVVTGTSAGSIVGNLFASGMSPDRLLEAEILGKTDLVDLTSTNG 120
orf137ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLLASGMSPDRLLEAEILGKTDLVDLTSTSG 120
50
orf137.pep FIKGAKLQNYINRKLGRMQIQFPIKFAA 149
orf137ng FIKGEKLQNYINRKVGGRRIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV 180

The complete length ORF137ng nucleotide sequence <SEQ ID 563> is:

55
1 ATGGAATA TGGTAACGTT TTCAAAAATC AGATCATTTT TGGCAATCGC
51 CGCCGCCGCG TTGCTTGCCG CCTGCGGTAC GGCAGGAAAC AATGCCGCC
101 GCAAGCCCGT GCAAAACGCC AAACCCGCCG CAGTGGTCGC TTGGCACTC
151 GGTGGCGCG CATCTAAGG ATTTGCCCAT ATAGGAATTG TTAAGGTTTT
201 GAAAGAAAAC GGTATTCCTG TGAAGGTGGT TACCGGCACA TCGGCAGGTT
251 CGATAGTCGG CAGCCTTTTG GCATCGGGTA TGTCGCCCGA CCGCCTCGAA
301 TTGGAAGCCG AGATTTTAGG TAAAACCGAT TTAGTCGATT TAACCTTGTC
351 CACCAAGTGT TTTATCAAAG GCGAAAAGCT GCAAAATTAC ATCAACCGAA
60 401 AAGTCGGCGG CAGGCAGATT CAGCAGTTTC CCATCAAATT TGCCGCCGTT
451 GCCACTGATT TTGAAACCGG CAAGGCCGTC GCTTTCAATC AAGGGAATGC

501 CGGGCAGGCG GTTCGTGCTT CCGCGGCCAT TCCCAATGTG TTCCAGCCAG
 551 TCATCATCGG CAGGCACAAA TATGTTGACG GCGGTCTGTC GCAGCCCGTG
 601 CCCGTCAGTG CCGCTCGGCG GCAGGGGGCG AATTCGTGA TTGCCGTCGA
 651 TATTTCCGCA CGTCCGAGCA AAAATGTCGG TCAAGGTTTC TTCTCTTATC
 701 TCGATCAGAC GCTGAACGTG ATGAGCGTTT CCGTGTTGCA AAACGAGTTG
 751 gggcAGGCGG ATGTGGTTAT CAAACCGCag gtTTGGATT TGGGTGCAGT
 801 CGGCGGATTG GATCAGAAAA AGCGCGCCAT CCGGTTGGGC GAGGAGGCAG
 851 CACGTGCCGC ATTGCCTGAA ATCAAACGCA AACTGGCGGC ATACCGTTAT
 901 TGA

10 This encodes a protein having amino acid sequence <SEQ ID 564>:

1 MENMVTFSKI RSFLAIAAAA LLAACGTAGN NAARKPVQTA KPAAVVALAL
 51 GGGASKGFAH IGIVKVLKEN GIPVKVVTGT SAGSIVGSL ASGMSPDRLE
 101 LEAEILGKTD LVDLTSTSG FIKGEKLQNY INRKVGGRQI QQFPIKFAAV
 151 ATDFETGKAV AFNQGNAGQA VRASAAIPNV FQPVIIGRHK YVDGGLSQPV
 15 201 PVSAARRQGA NFVIAVDISA RPSKNVQGF FSYLDQTLNV MSVSVLQNEL
 251 GQADVVIKPO VLDLGAVGGF DQKKRAIRLG EEAARAALPE IKRKLAAARY
 301 *

ORF137ng and ORF137-1 show 96.0% identity in 300 aa overlap:

20 orf137ng MENMVTFSKIRSFLAIAAAALLAACGTAGNNAARKPVQTAKPAAVVALALGGASKGFAH
 orf137-1 MENMVTFSKIRLLAIAAAALLAACGTAGNNAVRKPVQTAKPAAVVGLALGGASKGFAH
 orf137ng IGIVKVLKENGIPVKVVTGTSAGSIVGSLASGMSPDRLELEAEILGKTDLVDLTSTSG
 25 orf137-1 VGIKVLKENGIPVKVVTGTSAGSIVGSLFASGMSPDRLELEAEILGKTDLVDLTSTSG
 orf137ng FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
 orf137-1 FIKGEKLQNYINRKVGGRQIQQFPIKFAAVATDFETGKAVAFNQGNAGQAVRASAAIPNV
 30 orf137ng FQPVIIGRHKYVDGGLSQPVVSAARRQGANFVIAVDISARPSKNVQGGFFSYLDQTLNV
 orf137-1 FQPVIIGRHTYVDGGLSQPVVSAARRQGANFVIAVDISARPGKNISQGGFFSYLDQTLNV
 35 orf137ng MSVSVLQNELGQADVVIKPOVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY
 orf137 MSVSALQNELGQADVVIKPOVLDLGAVGGFDQKKRAIRLGEEAARAALPEIKRKLAAARY

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and
 40 *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 68

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 565>:

45 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA
 51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCTGcTG CCGCTTTCTT
 101 GTCTGCACAC GCTGGGAAAC CCGCTCGGAC ATCTGGCGTT TTACCTTTTA
 151 AAGGAAGACC GCGCGCGCAT CGTCGCmAT ATCGGCAGG CCGGTTTGAA
 201 CCCCAGCCCC AAAACGGTCA AAGCCGTTT TGCAGAAACG GCAAAAGGCG
 251 GTTTGAAGT TGCCCCCGCG TTTTCAGAA AACCGGAAGA CATAGAAACA
 50 301 ATGTTCAAAG CGGTACACGG CTGGAACAT GTGCAGCAGG CTTTGGACAA
 351 ACACGAAGG CTGCTATTC..

This corresponds to the amino acid sequence <SEQ ID 566; ORF138>:

1 MFRLQFRLFP PLRTAMHILL TALLKCLSL PLSClHTLGN RLGHlafYLL
 51 KEDRARIVAX MRQAGLNPD KTVKAVFAET AKGGLLAPA FFRKPEDIET
 55 101 MFKAVHGEH VQALDKHEG LLF

Further work revealed the complete nucleotide sequence <SEQ ID 567>:

```

      1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA
    51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGCTG CCGCTTTCCT
   101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA
   151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGGCAGG CGGGTTTGAA
   201 CCCCACCCCA AAAACGGTCA AAGCCGTTTT TGCGGAAACG GCAAAAGGCG
   251 GTTTGGAAC TGCCTCCGCG TTTTTCAGAA AACCAGGAAG CATAGAAACA
   301 ATGTTCAAAG CGGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGGACAA
   351 ACACGAAGGG CTGCTATTCA TCACGCCGCA CATCGGCAGC TACGATTGG
   401 GCGGACGCTA CATCAGCCAG CAGCTTCCGT TCCCCTGAC CGCCATGTAC
   451 AAACCGCCGA AAATCAAAGC GATAGACAAA ATCATGCAGG CGGGCAGGGT
   501 TCGCGGCAAA GGA AAAACCG CGCCTACCAG CATACAAGGG GTCAAAACAA
   551 TCATCAAAGC CCTGCGTTTC GCGGAAGCAA CCATCGTCCT GCCCGACCAC
   601 GTCCCTCCC CTCAAGAAGG CGGGGAAGGC GTATGGGTGG ATTTCTTCGG
   651 CAAACCTGCC TATACCATGA CGCTGGCGGC AAAATTGGCA CACGTCAAAG
   701 GCGTGAAAAC CCTGTTTTTC TGCTGCGAAC GCCTGCCTGG CGGACAAGGT
   751 TTCGATTGTC ACATCCGCCC CGTCCAAGGG GAATTGAACG GCGACAAAGC
   801 CCATGATGCC GCCGTGTTCA ACCGCAATGC CGAATATGG ATACGCCGTT
   851 TTCCGACGCA GTATCTGTTT ATGTACAACC GCTACAAAAT GCCGTAA

```

20 This corresponds to the amino acid sequence <SEQ ID 568; ORF138-1>:

```

      1 MFRLQFRLFP PLRTAMHILL TALLKCLSL PLSCSLHTLGN RLGLAFYLL
    51 KEDRARIVAN MRQAGLNPD KTVKAVFAET AKGGLELAPA FFRKPEDIET
   101 MFKAHVHWEH VQALDKHEG LLFITPHIGS YDLGGYISQ QLPFPLTAMY
   151 KPPKIKAIK IMQAGRVRGK GKTAPTSIOG VKQIIKALRS GEATIVLPDH
   201 VPSPEQEGEG VWVDFGKPA YMTLAAKLA HVKGVKTLFF CCERLPGGQG
   251 FDLHIRPVQG ELNGDKAHDA AVFNRAEYW IRRFPTQYLF MYNRYKMP*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF138 shows 99.2% identity over a 123aa overlap with an ORF (ORF138a) from strain A of *N.*

30 *meningitidis*:

```

      10      20      30      40      50      60
orfl38.pep MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCSLHTLGNRLGLAFYLLKEDRARIVAX
            |||
orfl38a    MFRLQFRLFPPLRTAMHILLTALLKCLSLPLSCSLHTLGNRLGLAFYLLKEDRARIVAN
      10      20      30      40      50      60

      70      80      90     100     110     120
orfl38.pep MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAHVHWEHVQALDKHEG
            |||
orfl38a    MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAHVHWEHVQALDKHEG
      70      80      90     100     110     120

orfl38.pep LLF
            |||
orfl38a    LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTSIOG
            130     140     150     160     170     180

```

The complete length ORF138a nucleotide sequence <SEQ ID 569> is:

```

      1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CCTTTGCGAA CCGCCATGCA
    51 CATCCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGCTG CCGCTTTCCT
   101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA
   151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGGCAGG CAGGCATGAA
   201 TCCCACCCCA AAAACGGTCA AAGCCGTTTT TGCGGAAACG GCAAAAGGCG
   251 GTTTGGAAC TGCCTCCGCG TTTTTCAGAA AACCAGGAAG CATAGAAACA
   301 ATGTTCAAAG CGGTACACGG CTGGGAACAT GTGCAGCAGG CTTTGGACAA
   351 ACACGAAGGG CTGCTATTCA TCACGCCGCA CATCGGCAGC TACGATTGG
   401 GCGGACGCTA CATCAGCCAG CAGCTTCCGT TCCCCTGAC CGCCATGTAC
   451 AAACCGCCGA AAATCAAAGC GATAGACAAA ATCATGCAGG CGGGCAGGGT
   501 TCGCGGCAAA GGA AAAACCG CGCCTACCAG CATACAAGGG GTCAAAACAA
   601 TCATCAAAGC CCTGCGTTTC GCGGAAGCAA CCATCGTCCT GCCCGACCAC

```

-327-

5
601 GTCCCTCCC CTCAAGAAGG CGGGGAAGGC GTATGGGTGG ATTTCTTCGG
651 CAAACCTGCC TATACCATGA CGCTGGCGGC AAAATTGGCA CACGTCAAAG
701 GCGTGAAAC CCTGTTTTTC TGCTGCGAAC GCCTGCCTGG CGGACAAGGT
751 TTCGATTTGC ACATCCGCCC CGTCCAAGGG GAATTGAACG GCGACAAAGC
801 CCATGATGCC GCCGTGTTCA ACCGCAATGC CGAATATTGG ATACGCCGTT
851 TTCCGACGCA GTATCTGTTT ATGTACAACC GCTACAAAT GCCGTAA

This encodes a protein having amino acid sequence <SEQ ID 570>:

10
1 MFRLOFRLFP PLRTAMHILL TALLKCLSL PLSCSLHTLGN RLGHAFYLL
51 KEDRARIVAN MRQAGLNPD KTVKAVFAET AKGGLELAPA FFRKPEDIET
101 MFKAHVHWEH VQALDKHEG LLFITPHIGS YDLGGYISO QLPFLTAMY
151 KPPKIKAIK IMQAGRVRGK GKTAPTSIQG VKQIKALRS GEATIVLPDH
201 VPSPQEGGEG VWVDFGKPA YTMTLAAKLA HVKGVKTLFF CCERLPGGQG
251 FDLHIRPVQG ELNGDKAHD AVFNRAEYW IRRFPTQYLE MYNRYKMP*

ORF138a and ORF138-1 show 99.7% identity over a 298aa overlap:

15 orf138a.pep MFRLOFRLFPPLRTAMHILLTALLKCLSLPLSCSLHTLGNRLGHAFYLLKEDRARIVAN
orf138-1 MFRLOFRLFPPLRTAMHILLTALLKCLSLPLSCSLHTLGNRLGHAFYLLKEDRARIVAN
20 orf138a.pep MRQAGMNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAHVHWEHVQALDKHEG
orf138-1 MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAHVHWEHVQALDKHEG
25 orf138a.pep LLFITPHIGSYDLGGYISOQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTSIQG
orf138-1 LLFITPHIGSYDLGGYISOQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTSIQG
30 orf138a.pep VKQIKALRSGEATIVLPDHVPSPQEGGEGVWVDFGKPAYTMTLAAKLAHVKGVKTLFF
orf138-1 VKQIKALRSGEATIVLPDHVPSPQEGGEGVWVDFGKPAYTMTLAAKLAHVKGVKTLFF
35 orf138a.pep CCERLPGGQGFDLHIRPVQGEELNGDKAHDAAVFNRAEYWIRRFPTQYLFMYNRYKMP
orf138-1 CCERLPGGQGFDLHIRPVQGEELNGDKAHDAAVFNRAEYWIRRFPTQYLFMYNRYKMP

35 Homology with a predicted ORF from *N.gonorrhoeae*

ORF138 shows 94.3% identity over a 123aa overlap with a predicted ORF (ORF138ng) from

N.gonorrhoeae:

40 orf138.pep MFRLOFRLFPPLRTAMHILLTALLKCLSLPLSCSLHTLGNRLGHAFYLLKEDRARIVAX 60
orf138ng MFRLOFRLFPPLRTAMHILLTALLKCLSLSLSCSLHTLGNRLGHAFYLLKEDRARIVAN 60
orf138.pep MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAHVHWEHVQALDKHEG 120
orf138ng MRQAGLNPDQTVKAVFAETAKGGLAPAFFKKPEDIETMFKAHVHWEHVQALDKGEG 120
45 orf138.pep LLF 123
orf138ng LLFITPHIGSYDLGGYISOQLPFHLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTGIQG 180

The complete length ORF138ng nucleotide sequence <SEQ ID 571> is:

50 1 ATGTTTCGTT TACAATTCAG GCTGTTTCCC CTTTGGCGAA CCGCCATGCA
51 CATCTGTTG ACCGCCCTGC TCAAATGCCT CTCCCTGCTG TCGCTTTCCT
101 GTCTGCACAC GCTGGGAAAC CGGCTCGGAC ATCTGGCGTT TTACCTTTTA
151 AAGGAAGACC GCGCGCGCAT CGTCGCCAAT ATGCGGCAGG CGGGTTTGAA
201 CCCCAGACAG CAGACGGTCA AAGCCGTTT TGCGGAAACG GCAAATGCG
55 251 GTTTGGAAGT TGCCCCCGCG TTTTCAAAA AACCGGAAGA CATCGAAACA
301 ATGTTCAAAG CGGTACACGG CTGGGAACAC GTGCAGCAGG CTTTGGACAA
351 GGGCGAAGGG CTGCTGTTCA TCACGCCGCA CATCGGCAGC TACGATTGCG
401 GCGGACGCTA CATCAGCCAG CAGCTTCCGT TCCACCTGAC CGCATGTAC
451 AAGCCGCCGA AAATCAAAGC GATAGACAAA ATCATGCAGG CGGGCGAGGT
60 501 GCGCGGCAAA GGCAAAACcg cggccaccgg catACAAGGG GTCAAACAAA
551 tcatcaAGGC CCTGCGCGCG GCGAGGCAA CCATcATCCT GCCCGACCAC

-328-

5
601 GTCCTTCTC CGCAGGAagg cggCGGCGTG TGGGCGGATT TTTTCGGCAA
651 ACCTGCATAc accATGACAC TGGCGGCAAA ATTGGCACAC GTCAAAGGCG
701 TGAACCCCT GTTTTCTGC TGGCAACGCC TGCCCGACGG ACAAGGCTTC
751 GTGTTGCACA TCCGCCCCGT CCAAGGGGAA TTGAACGGCA ACAAGCCCA
801 CGATGCCGCC GTGTTCAACC GCAATACCGA ATATTGGATA CGCCGTTTTC
851 CGACGCAGTA TCTGTTTATG TACAACCGCT ATAAACGCC GTAA

This encodes a protein having amino acid sequence <SEQ ID 572>:

10
1 MFRLQFRLFP PLRTAMHILL TALLKCLSL LSLCLHTLGN RLGHlafYLL
51 KEDRARIVAN MRQAGLNPD TQTKAVFAET AKCGLELAPA FFKKPEDTET
101 MFKAVHGW EHVQQALDKGEG LLFITPHIGS YDLGGYISQ QLPFHLTAMY
151 KPPKIKAIK IMQAGRVRGK GKTAPTGIQ VQKIIKALRA GEATIILPDH
201 VPSPQEGGV WADFFGKPAY TMTLAAKLAH VKGVKTLFFC CERLPDQGF
251 VLHIRPVQGE LNGNKAHDAA VFNRTYEWI RRFPTQYLFM YNRYKTP*

ORF138ng and ORF138-1 show 94.3% identity over 299aa overlap:

15 orf138-1.pep MFRLQFRLFPPLRTAMHILLTALLKCLSL LSLCLHTLGNRLGHlafYLLKEDRARIVAN
orf138ng MFRLQFRLFPPLRTAMHILLTALLKCLSL LSLCLHTLGNRLGHlafYLLKEDRARIVAN
20 orf138-1.pep MRQAGLNPDPKTVKAVFAETAKGGLELAPAFFRKPEDIETMFKAVHGW EHVQQALDKHGE
orf138ng MRQAGLNPD TQTKAVFAETAKCGLELAPAFFRKPEDIETMFKAVHGW EHVQQALDKGEG
25 orf138-1.pep LLFITPHIGSYDLGGYISQQLPFPLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTSGI
orf138ng LLFITPHIGSYDLGGYISQQLPFHLTAMYKPPKIKAIKIMQAGRVRGKGKTAPTGIQ
30 orf138-1.pep VKQIIKALRSGEATIVLPDHPVSPQEGGVWVDFFGKPAYTMTLAAKLAHVKGVTLLFF
orf138ng VKQIIKALRAGEATIILPDHPVSPQEGG-GVWADFFGKPAYTMTLAAKLAHVKGVTLLFF
orf138-1.pep CCERLPGQGFDLHIRPVQGE LINGDKAHDAAVFNRTYEWIRRFPTQYLFMYNRYKMP
orf138ng CCERLPGQGFVLHIRPVQGE LINGNKAHDAAVFNRTYEWIRRFPTQYLFMYNRYKTP

In addition, ORF138ng is homologous to htrB protein from *Pseudomonas fluorescens*:

35 gnl|PID|e334283 (Y14568) htrB [*Pseudomonas fluorescens*] Length = 253
Score = 80.8 bits (196), Expect = 9e-15
Identities = 49/151 (32%), Positives = 79/151 (51%), Gaps = 6/151 (3%)
40 Query: 101 MFKAVHGW EHVQQALDKGEGLLFITPHIGSYD-LGGYISQQLPFHLTAMYKPPKIKAIK 159
+ + V G E +++AL G+G+ IT H+G+++ L Y SQ P Y+PPK+KA+D
Sbjct: 94 LVREVEGLEVLKEALASGKGVGITSHLGNWEVLNHFYCSQCKPI---IFYRPPKLKA VD 150
Query: 160 KIMQAGRVRGKGKTAPTGIQGVKQIIKALRAGEATIILPDHPVSPQEGGVWADFFGKPA 219
++++ RV+ K A + +G+ +IK +R G I D P P E G++ FF A
45 Sbjct: 151 ELLRKQVRQLGNKVAASKEGILSVIKEVRKGGQVGIPAD--PEPAESAGIFVFFATQA 208
Query: 220 YMTLAAKLAHVKGVTLLFFCCERLPDQGF 250
T + +F RLPDG G+
Sbjct: 209 LTSKFVPNMLAGGKAVGVFLHALRLPDGSGY 239

50 Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF138-1 (57kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above.

The products of protein expression and purification were analyzed by SDS-PAGE. Figure 14A

55 shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis

(Figure 14B). These experiments confirm that ORF138-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 69

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 573>:-

```

5      1  ..GCGTGGTCGG CCGGCCAATC GTGGCGTGTG TTAATGGAAA GTGAAACGTG
51     GCATGCGGTG TGGAAACTT  TCGCTTCTC  GCGGCGGCGG GTGTATGCGG
101    CAGCGGTTT  GGGTGTGGTG TATGCGGCGC GCGCGCGGCG GTGCGCGTGG
151    ATGCGCGGGC TGATGTTT  TTA  GCGGTTTATG GTGTGCGCCG TTTGTGTTTC
201    GCGGCGCGTG CTGCTGCTT  T  ATCCGCAGTG GACGGCTTCG TTGCCGTGTC
10     251   TGCTGGCGAT  GTATGCGCTG CTGGCGTATC CGTTTGTGGC AAAAGATGTT
301    TTATCAGCCT GGGATGCACT GCCGCCGGAT TACGGCAGGG CCGCGCGCGG
351    TTTGGGTGCA AACGGCTTTC AGACGGCATG CCGCATCACG TTCCCCCTCT
401    TGAACCGGCG GTTGCGGCGC GGTCTGACTT TGGCGGCGGC AACCTGCGTG
451    GCGCAATTG  CCGCGACATT GTTCTGTCTG CGTCCGGAAT GGCAGACGCT
15     501    GACGACTT  T  ATTTATGCCT ATTTGGGACG CCGCGGTGAG GATAATTACG
551    CGCGGGCGAT GGTGCTG..

```

This corresponds to the amino acid sequence <SEQ ID 574; ORF139>:

```

1      1  ..AWSAGESWRV LMESETWHAV WNTLRFSAAA VYAAAVLGVV YAAPARRSAW
51     MRGLMFXPEM VSPVCVSAGV LLLYPQWTAS LPLLLAMYAL LAYPFVAKDV
20     101    LSAWDALPPD YGRAAGLGA NGFQTACRIT FPLLPALRR GLTLAAATCV
151    GEFAATLFLS RPEWQTLTTL IYAYLGRAGE DNYARAMVL..

```

Further work revealed the complete nucleotide sequence <SEQ ID 575>:

```

1      1  ATGGATGGAC GCGTGGGT  GGTATGGGGT GCTTTTGCCC TGCTGCCTTC
25     51     GGCTTTTTTG GCGGTAATGG TCGTTGCGCC TTTGTGGGCG GTGGCGGCGT
101    ATGACGGTTT GCGGTGGCGC GCGGTGCTGT CGGATGCCTA TATGCTCAAA
151    CGTTTGGCGT GGACGGTATT TCAGGCAGCG GCAACCTGTG TGCTGGTGCT
201    GCCTTTGGGC GTGCCTGTCT CGTGGGTGCT GCGCGGCGTG GCGTTTCCGG
251    GCGGGGCTTT GGTGCTGCGC CTGCTGATGC TGCCTTTTGT GATGCCACG
301    TTGGTGGCGG GCGTGGGCGT GCTGGCCCTG TTCGGGGCGG ACGGGCTGTT
10     351    GTGGCGCGGC AGGCAGGATA CGCCGTATCT GTTGTGTAC GGCAATGTG
401    TTTTCAACCT TCCTGTGTTG GTCAGGGCGG CGTATCAGGG GTTTGTGCAA
451    GTGCCTGCGG CACGGCTTCA GACGGCACGG ACCTTGGGCG CGGGGGCGTG
501    GCGGCGGTTT TGGGACATTG AAATGCCCGT TTGCGGCCG TGGCTTGCCG
551    GCGGCGTGTG CCTTGTCTTT CTGTATTGTT TTCCGGGTT CCGGCTGGCG
35     601    CTGCTGCTGG GCGGCAGCCG TTATGCCACG GTCGAAGTGG AAATTTACCA
651    GTTGGTCATG TTCGAACCTG ATATGGCGGT TGCTTCGGTG CTGGTGTGGC
701    TGGTGTGGG  GGTAAACGGC GCGGCAGGGT TGCTGTATGC GTGGTTCGGC
751    AGGCGCGCGG TTTCGGATAA GGCGGTTTCC CTGTGTATGC CGTCGCGGCC
801    GCAGTCGGTC GGGGAATATG TGCTGCTGGC GTTTCGCGCG GCGGTGTTGT
40     851    CTGTGTGCTG CCTGTTTCTT TTGTTGGCAA TTGTTGTGAA AGCGTGGTCTG
901    GCCGCGGAAT CGTGGCGTGT GTTAATGGAA AGTGAACGT GGCAGGCGGT
951    GTGGAATACT TTGCGCTTCT CGGCGGCGGC GGTGTATGCG GCGGCGGTTT
1001   TGGGTGTGGT GTATGCGGCG GCGGCGCGGC GGTGCGCGTG GATGCGCGGG
1051   CTGATGTTTT TGCCGTTTAT GGTGTGCGCG GTTGTGTTT CCGCGGCGGT
45     1101  GCTGCTGCTT TATCCGAGT GGACGGCTTC GTTCCGTTG CTGCTGGCGA
1151   TGTATGCGCT GCTGGCGTAT CCGTTTGTGG CAAAAGATGT TTTATCAGCC
1201   TGGGATGCAC TGCCGCCGGA TTACGGCAGG GCGGCGGCGG GTTGGGTGTC
1251   AAACGGCTTT CAGACGGCAT GCCGCATCAC GTTCCCCCTC TTGAAACCGG
1301   CGTTGCGGCG CGGTCTGACT TTGGCGGCGG CAACCTGCGT GGGCGAATTT
50     1351   GCGGCGACAT TGTTTCTGTC GCGTCCGGAA TGGCAGACGC TGACGACTTT
1401   GATTTATGCC TATTTGGGAC GCGCGGGTGA GGATAATTAC GCGCGGGCGA
1451   TGGTGCTGAC ATTGCTGTTG GCGGCGTTCG CGCTGGGTAT TTTCTGCTG
1501   TTGGACGGCG GCGAAGGCGG AAAACAGACG GAAACGTTAT AA

```

This corresponds to the amino acid sequence <SEQ ID 576; ORF139-1>:

```

55     1      1  MDGRRWVWVG AFALLPSAFL AVMVVAPLWA VAAVDGLAWR AVLSDAYMLK
51     RLAWTVFQAA ATCVLVPLG VPAWVLAARL AFPGRALVLR LLMLPFVMP
101    LVAGVGVLA  FGADGLLWRG RQDTPYLLLY GNVFFNLPVL VRAAYQGFVQ
151    VPAARLQ  TAR  TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA

```


5

Homology with a predicted ORF from *N.meningitidis* (strain A)

10

15

30

35

40

45

50

55

60

1401 GATTTATGCC TATNTGGGAC GCGCGGGTGA NGATAATTAC GCGCGGGCGA
 1451 TGGTGCTGAC ATTGCTGTTG GCGGCGTTCG CGCTGGGTAT NTTCTGCTG
 1501 TTGGACGGCG GCGAAGGCGG AAAACGGACG GAAACGTTAT AA

This encodes a protein having amino acid sequence <SEQ ID 578>:

5 1 MDGRRWAVWG AFALLPSAFL AAMVVAPLWA VAAYDGLAWR AVLSDAYMLK
 51 RLAWTVFQAA ATCVLVPLG VPAWVLARL AFPGRALVLR LLMLPFVMP
 101 LVAGVGV LAL FGADGLXWRG WQDTPYLLLY GNVFFXLPVL VRAAYQG FVQ
 151 VPAARLOTAX TLGAGAWRRF WDIEMPVLRP WLAGGVCLVF LYCFSGFGLA
 201 LLLGGSRYAT VEVEIYQLVM FELDMAVASV LVWLXGVTA AAGLLYAWFG
 10 251 RRAVSDKAVS PVMPSPQSV GEYVLLAF AA AVXSVCCLFX LLAIIVKAW
 301 AGESWRV LME SETWQAVWNT XRFSA AAVYA AAVLG VVYAA AARRSAWMRG
 351 LMFLPFMVSP VCVSAGV LLL XPQWTAS LPL LLAMYALLAY PFVAKDVLSA
 401 XDALPPDYGR AAAGLGANGF QTACRITFPL LKPALRRGLT LAAATCVGEF
 451 AATLFXSRXE WQTLTLLIYA YXGRAGXDNY ARAMVLTLL L AAFALGXELL
 15 501 LDGGE GGGKRT ETL*

ORF139a and ORF139-1 show 96.5% homology over a 514aa overlap:

orf139a.pep MDGRRWAVWGAFALLPSAFLAAMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
 orf139-1 MDGRRWVWGAFALLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
 20 orf139a.pep ATCVLVPLGVPVAVWLARLAFPGRALVLRLLMLPFVMP TLVAGVGV LALFGADGLXWRG
 orf139-1 ATCVLVPLGVPVAVWLARLAFPGRALVLRLLMLPFVMP TLVAGVGV LALFGADGLXWRG
 25 orf139a.pep WQDTPYLLLYGNVFFXLPVLVRAAYQGFVQVPAARLOTAXTLGAGAWRRFWDIEMPVLRP
 orf139-1 RQDTPYLLLYGNVFFNLPVLVRAAYQGFVQVPAARLOTARTLGAGAWRRFWDIEMPVLRP
 30 orf139a.pep WLAGGVCLVFLYCFSGFGLALLGGSRYATVEVEIYQLVMFELDMAVASVLWLXGVTA
 orf139-1 WLAGGVCLVFLYCFSGFGLALLGGSRYATVEVEIYQLVMFELDMAVASVLWLXGVTA
 35 orf139a.pep AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAF AA AVXSVCCLFXLLAIIVKAW
 orf139-1 AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAF AA AVXSVCCLFXLLAIIVKAW
 40 orf139a.pep AGESWRV LMESETWQAVWNTXRFSA AAVYA AAVLG VVYAAAARRSAWMRGLMFLPFMVSP
 orf139-1 AGESWRV LMESETWQAVWNTLRFSA AAVYA AAVLG VVYAAAARRSAWMRGLMFLPFMVSP
 45 orf139a.pep VCVSAGV LLLXPQWTASLPLLLAMYALLAYPFVAKDVLSAXDALPPDYGRAAAGLGANGF
 orf139-1 VCVSAGV LLLXPQWTASLPLLLAMYALLAYPFVAKDVLSA W DALPPDYGRAAAGLGANGF
 50 orf139a.pep QTACRITFPL LKPALRRGLT LAAATCVGEFAATLFXSRXEWQTLTLLIYAYXGRAGXDNY
 orf139-1 QTACRITFPL LKPALRRGLT LAAATCVGEFAATLFLSRPEWQTLTLLIYAYLGRAGEDNY
 orf139a.pep ARAMVLTLL L AAFALGXELL LDGGE GGGKRTETLX
 orf139-1 ARAMVLTLL L AAFALGIFLL LDGGE GGGKQTETLX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF139 shows 95.2% identity over a 189aa overlap with a predicted ORF (ORF139ng) from

55 *N.gonorrhoeae*:

orf139.pep AWSAGESWRV LMESETWQAVWNTLRFSA A A 30
 orf139ng QSVGEYVLLAFSAVAVLSVCCLFPLSAIVVKAWSAGESRRV LMESETWQAVWNTLRFSA A A 327
 60 orf139.pep VYAAAVLG VVYAAAPARRSAWMRGLMFXPFMVSPVCVSAGV LLLXPQWTASLPLLLAMYAL 90
 orf139ng VFAAAVLG VVYAAAARRLVMMRGLVFLPFMVSPVCVSAGV LLLXPQWTASLPLLLAMYAL 387

10	1	MDGRCWAVRG	AFSLLPSAFL	AVMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTVFQAA	ATCVLVLPGL	VPVAVVLARL	AFPGRALVLR	LLMLPFVMP
	101	LVAGVGLAL	FGADGLLWRG	RQDTPYLLLY	GNVFFNLPLV	VRAAYQGFAQ
	151	VPAARLQTAR	FLGAGAWRPF	WDIEMPVLRP	WLAGGVCLVF	LYCFSQGFGL
	201	LLGGSSRYAT	VEVEIYQLVM	FELDMAGASA	LVWLVLGVTA	AAGLLYAWFG
15	251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFSV	AVLSVCCCLF	LSAIVVKVKS
	301	AGESRRVLME	SETWQAVWNT	LRFSAAAVFA	AAVLGVVYAA	AARRLVWMRG
	351	LVFLPFMVSP	VCVSAGVLLL	YPGWTASLPL	LLAMYALLAY	PFVAKDVLSA
	401	WDALPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAAATCVGEF
	451	AATLFLSRPE	WQTLTTLIYA	YLGRAGEDNY	ARAMVLTLLL	SAFAVCIFLL
20	501	LDNEGGKRT	ETL*			

	1	ATGGATGGAC	GGTGTGGGC	GGTACGGGGT	GCTTTTTCCC	TGCTGCCTTC
	51	GGCTTTTTTG	GCGTAATGG	TCGTTGCGCC	TTTGTGGGCG	GTGGCGGCGT
25	101	ATGACGGTTT	GGCGTGGCG	GCGGTGCTGT	CGGATGCCTA	TATGCTCAAA
	151	CGTTTGCGCT	GGACGCTGTT	TCAGCGGCG	GCAACCTGTG	TGCTGGTGCT
	201	GCCTTTGGGC	GTGCGTGTGC	CTGGGGTGCT	GGCGCGGCTG	GC GTTCCCGG
	251	GGCGGGCTTT	GGTGCTGCGC	CTGCTGATGC	TGCCGTTTGT	GATGCCACGC
	301	CTGGTGGCGG	GCGTGGGCGT	GCTGGCTCTG	TTCGGGGCGG	ACGGGCTGTT
30	351	GTGGCGCGCG	CGGCAGGATA	CGCCGTATCT	GTTGTTGTAC	GGCAATGTGT
	401	TTTTCAACCT	GCCCGTGTTG	GTCAGGGCGG	CGTATCAGGG	GTTTGCTCAA
	451	GTGCCTGCGG	CACGGCTTCA	GACGGCACGG	ACGTTGGGCG	CGGGGGCGTG
	501	GCGGCGGTTT	TGGGACATTG	AAATGCCCGT	TTTGGCGCCC	TGGCTTCCCG
	551	GCGGCGGTGT	CCTTGCTTTC	CTGTATTGTT	TTTCGGGGTT	CGGGCTGGCA
35	601	TTGCTGTTGG	GCGGCAGCCG	TTATGCCACG	GTCGAAGTGG	AAATTTACCA
	651	GTTTGTTATG	TTCGAACTCG	ATATGGCGGG	GGCTTCGGCG	CTGGTGTGGC
	701	TGGTGTTGGG	GGAACCGCG	GCGGCAGGGT	TGCTGTATGC	GTGGTTCGGC
	751	AGGCGCGCGG	TTTCGGATAA	GCGGTTTCC	CCCGTGATGC	CGTCCGCCGC
	801	GCAATCGGTG	GGGGAATATG	TATTGCTGGC	ATTTTCGGTG	GCGGTGTTGT
40	851	CCGTGTGCTG	CCTGTTTCTC	TTGTCCGGAA	TTGTTGTGAA	AGCGTGGTCT
	901	GCCGCGCAAT	CGCGCGGTGT	GTTAATGGAA	AGTGAACAGT	GGCAGGACGT
	951	TGGGAATACT	ttGCGCTTTT	CGCGCGCGGC	GGTGTTTGCG	GCGGCGGTTT
	1001	TGGGTGTGGT	GTATGCGGCG	GCGGCGCGGC	GGCTGGTGTG	GATGCGCGGA
	1051	CTGGTGTTTT	TACCGTTTAT	GGTGTCGCCG	GTTTGCTGTT	CGCGGGGCGT
45	1101	GCTGCTGCTT	TATCCGGGGT	GGACGGCTTC	GTTACCGCTG	GTTGCGGCGA
	1151	TGTATGCGCT	GCTGGCGTAT	CCGTTTGTGG	CAAAAAGATG	TTTATCGGCC
	1201	TGGGATGCAC	TGCCGCCGGA	TTACGGCAGG	GCGGCGGCAG	TTTGGGCGCG
	1251	AAACGGCTTT	CAGACGGCAT	GCCGTATCAC	GTCCCCCTC	TTGAAACCGG
	1301	CGTTGCGGCG	GCGTCTGACT	TGCGCGCGCG	CGACGTGTGT	GGCGGAATTT
	1351	GCGGCAACCT	TGTTCTCTGC	GCGTCCGGAA	TGGCAGACGT	TGACGACTTT
50	1401	GATTTATGCC	TATTTGGGGC	GTGCGGGTGA	GGACAATTAT	GCGCGGGCAA
	1451	TGGTGTTGAC	ATTGCTGTTG	TCGGCATTGT	CGGTGTGCAT	TTTCTTGCTG
	1501	TGGACAACCG	aaaACGGACG	aaaACGGACG	GAAACGTTAT	AA

55	1	MDGRCWAVRG	AFSLLPSAFL	AVMVVAPLWA	VAAYDGLAWR	AVLSDAYMLK
	51	RLAWTFVQAA	ATCVLVLPFG	VPVAVVLARL	AFPGRALVLR	LLMLPFVMPT
	101	<u>LVAGVGVIAL</u>	<u>FGADGLLWRG</u>	<u>RQDTPYLLLY</u>	<u>GNVFFNLPLV</u>	<u>VRAAYGQFAQ</u>
	151	VPAARLQOTAR	TLGAGWRRF	WDIEMPVLRP	WLAGGVCVLF	LYCFSGGFLA
	201	LLGGSSRYAT	VEVEIYQLVM	FELDMAGASA	<u>LVWLVLGVTA</u>	<u>AAGLLYAWFG</u>
60	251	RRAVSDKAVS	PVMPSPQSV	GEYVLLAFSV	<u>AVLSVCCFLP</u>	<u>LSAIVVKWAS</u>
	301	AGESRRVLME	SETWOQVWNT	FLRSAAAVFA	<u>AAVLGVVYAA</u>	<u>AARRLVWMRG</u>
	351	LVFLPFMVSP	<u>VCVSAGVLLL</u>	<u>YPGWTASLPL</u>	<u>LLAMYALLAY</u>	<u>PFVAKDVLSA</u>
	401	WDAIPPDYGR	AAAGLGANGF	QTACRITFPL	LKPALRRGLT	LAATCVGEF
	451	AATLFLSRPE	WQTLTTLIYA	YLGRAGEDNY	<u>ARAMVLTLLL</u>	<u>SAFAVCIPLL</u>
	501	LONGEGKRT	ETL*			

ORF139ng-1 and ORF139-1 show 95.9% identity over 513aa overlap:

```

5      orf139ng      MDGRCWAVRGAFSLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA
      orf139-1      MDGRRWVWVGAFALLPSAFLAVMVVAPLWAVAAYDGLAWRAVLSDAYMLKRLAWTVFQAA

10     orf139ng      ATCVLVLPPLGVPVAWVLARLAFPGRALVLRLLMLPFVMPPLVAGVGVLAIFGADGLLWRG
      orf139-1      ATCVLVLPPLGVPVAWVLARLAFPGRALVLRLLMLPFVMPPLVAGVGVLAIFGADGLLWRG

15     orf139ng      RQDTPYLLLYGNVFFNLPLVLRAAAYQGFAQVPAARLQTARTLGAGAWRRFWDIEMPVLRP
      orf139-1      RQDTPYLLLYGNVFFNLPLVLRAAAYQGFVQVPAARLQTARTLGAGAWRRFWDIEMPVLRP

20     orf139ng      WLAGGVCLVFLYCFSGFGLALLLGGSRATVEVEIYQLVMFELDMAGASALVWLVLGVTA
      orf139-1      WLAGGVCLVFLYCFSGFGLALLLGGSRATVEVEIYQLVMFELDMASVSVLWLVLGVTA

25     orf139ng      AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFSAVLSVCCFLPLSAIVVKAWS
      orf139-1      AAGLLYAWFGRRRAVSDKAVSPVMPSPQSVGEYVLLAFSAVLSVCCFLPLSAIVVKAWS

30     orf139ng      AGESRRVLMESETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP
      orf139-1      AGESRRVLMESETWQAVWNTLRFSAAAVFAAAVLGVVYAAAARRLVWMRGLVFLPFMVSP

35     orf139ng      VCVSAGVLLLYPGWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAGLGANGF
      orf139-1      VCVSAGVLLLYPGWTASLPLLLAMYALLAYPFVAKDVLSAWDALPPDYGRAAGLGANGF

40     orf139ng      QTACRITFPLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY
      orf139-1      QTACRITFPLKPALRRGLTLAAATCVGEFAATLFLSRPEWQTLTTLIYAYLGRAGEDNY

50     orf139ng      ARAMVLTLLLSAFVAVCIFLLLDNGEGGKRTETL
      orf139-1      ARAMVLTLLLSAFVAVCIFLLLDNGEGGKRTETL

```

Based on the presence of a predicted binding-protein-dependent transport systems inner membrane component signature (underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 70

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 583>:

```

45      1  ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTT
      51  GGCGGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAGA TTCCGCATCC
      101  ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC
      151  GGTTCGCCCA CAGGCAGCAT TGTCAAAGAC ATACTGGTCA AAAACTTCGG
      201  CGGCACGCTC GGCGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG
      251  AACGTTGGT C...

```

This corresponds to the amino acid sequence <SEQ ID 584; ORF140>:

```

50      1  MDGWTQTLQA QTLIGISAAA IILILILIVR FRIHALLTLV IVSLLTALAT
      51  GLPTGSIVKD ILVKNFGGTL GGVALLVLGL AMLERLV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 585>:

```

55      1  ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTT
      51  GGCGGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAAA TTCCGCATCC
      101  ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC
      151  GGTTCGCCCA CAGGCAGCAT TGTCAAAGAC ATACTGGTCA AAAACTTCGG

```

201 CGGCACGCTC GCGCGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG
 251 GACGTTTGGT CGAAACATCC GCGCGCGCAC AGTCGCTGGC GGACGCGCTG
 301 ATCCGGATGT TCGGCGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC
 401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCCC
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCTGCCC
 501 GCCCCATCCG GGCCCGATTG CCGCTTCCGA ATTTTACGGC GCGAACATCG
 551 GCCAAGTTT GATTTTGGGT CTGCCGACCG CCTTCATCAC ATGGTATTTT
 601 AGCGGCTATA TGCTCGGCAA AGTGTTGGGG CGCACCATCC ATGTTCCCGT
 651 TCCCGAAGTG CTCAGCGGCG GCACGCAAGA CAACGACCTG CCGAAAGAAC
 701 CTGCCAAAGC AGGAACGGTC GTCGCCATCA TGCTGATTCC CATGCTGCTG
 751 ATTTTCTTGA ATACCGGCGT ATCGGCCCTC ATCAGCGAAA AACTCGTAAG
 801 TGCGGACGAA ACCTGGGTTC AGACGGCAA AATAATCGGT TCGACACCGA
 851 TCGCCCTTCT GATTTCCGTA TTGGTCGCAC TGTTTGTCTT GGGACGCAAA
 901 CGCGGCGAAA GCGGCAGCGC GTTGGAAAAA ACCGTGGACG GCGCACTCGC
 951 CCGCTCTGT TCCGTGATTC TGATTACCGG CGCGGGCGGT ATGTTCCGGC
 1001 GCGTTTTGCG CGCTTCCGGC ATCGGCAAGG CACTCGCCGA CAGCATGGCG
 1051 GATTTGGGCA TTCCCGTCTT TTTGGGCTGT TTCCTTGTG CCTTGGCACT
 1101 GCGTATCGCG CAAGGTTTCGG CAACCGTCGC CCTGACCACC GCCGCCGCGC
 1151 TGATGGCTCC TGCCGTTGCC GCCGCCGGCT TTACCGACTG GCAGCTCGCC
 1201 TGTATCGTAT TGGCAACGGC GGCAGGTTTC GTCCGTTGCA GCCACTTCAA
 1251 CGACTCCGGC TTCTGGCTGG TCGGCCGTCT CTTGGACATG GACGTACCGA
 1301 CCACGCTGAA AACCTGGACG GTCAACCAAA CCCTCATCGC ACTCATGGC
 1351 TTTGCCTTGT CCGCACTGCT GTTCGCCATC GTCTGA

25 This corresponds to the amino acid sequence <SEQ ID 586; ORF140-1>:

1 MDGWTQTLISA QTLGISAAA IILILILIVK FRIHALTLV IVSLLTALAT
 51 GLPTGSIVND ILVKNFGGTL GGVALLVGLG AMLGRLVETS GGAQSLADAL
 101 IRMFGEKRAP FALGVASLIF GFPIFFDAGL IVMLPIVFAT ARRMQDVLV
 151 FALASIGAFS VMHVFLPPHP GPIAASEFYG ANIGQVLILG LPTAFITWYF
 201 SGYMLGKVLG RTIHVPVPEL LSGGTQDNDL PKEPAKAGTV VAIMLIPMLL
 251 IFLNTGVSAL ISEKLVSAD E TWVQTAKIIG STPIALLISV LVALFVLGRK
 301 RGESESALEK TVDGLAPVC SVILITGAGG MFGGVLRASG IGKALADSMA
 351 DLGIPVLLGC FLVALALRIA QGSATVALTT AAALMAPAVA AAGFTDWQLA
 401 CIVLATAAGS VGCSEFNDG FWLVGRLLDM DVPTTLKWT VNOTLIALIG
 451 FALSALLFAI V*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF140 shows 95.4% identity over a 87aa overlap with an ORF (ORF140a) from strain A of *N. meningitidis*:

40		10	20	30	40	50	60
orf140.pep		MDGWTQTLISAQTLGISAAAIIILILILIVRFRIHALTLVIVSLLTALATGLPTGSIVKD					
orf140a		MDGWTQTLISAQTLGISAAAIIILILILIVKFRHALTLVIVSLLTALATGLPTGSIVND					
		10	20	30	40	50	60
45		70	80				
orf140.pep		ILVKNFGGTLGGVALLVGLGAMLERLV					
orf140a		VLVKNFGGTLGGVALLVGLGAMLRVETSSGGAQSLADALIRMFGEKRAPFALGVASLIF					
50		70	80	90	100	110	120

The complete length ORF140a nucleotide sequence <SEQ ID 587> is:

1 ATGGACGGCT GGACACAGAC GCTGTCCGCG CAAACCCTGT TGGGCATTTC
 51 GCGGCGGCA ATCATCCTCA TTCTGATTTT AATCGTCAA TCCGCGATCC
 101 ACGCGCTGCT GACACTGGTC ATCGTCAGCC TGCTGACGGC TTTGGCAACC
 151 GGTGTGCCCA CAGGCAGCAT TGTCAACGAC GTACTGGTCA AAAACTTCGG
 201 CGGCACGCTC GCGGCGTGG CGCTTCTGGT CGGCCTGGGC GCGATGCTCG
 251 GACGTTTGGT CGAAACATCC GCGGCGCAC AGTCGCTGGC GGACGCGCTG
 301 ATCCGGATGT TCGGCGAAAA ACGCGCACCG TTCGCGCTGG GCGTTGCCTC
 351 GCTGATTTTC GGCTTCCCGA TTTTCTTCGA TGCCGGACTA ATCGTCATGC
 401 TGCCCATCGT GTTCGCCACC GCACGGCGCA TGAAACAGGA CGTACTGCCC
 451 TTCGCGCTTG CCTCCATCGG CGCATTTTCC GTCATGCACG TCTTCTGCCC

501 G C C C C A T C C G G G C C G A T T G C C G T T C C G A A T T T A C G G C G C G A A C A T C G
 551 G C C A A G T T T T G A T T T T G G G T C T G C C G A C C G C C T T C A T C A C A T G G T A T T T C
 601 A G C G G C T A T A T G C T C G G C A A A G T G T T G G G G C G C A C C A T C C A T G T T C C C G T
 651 T C C C G A A C T G C T C A G C G G C G C A C G C A A G A C A A C G A C C T G C G A A A G A A C
 701 C T G C C A A A G C A G G A A C G G T C G T C C C A T C A T G T G A T T C C A T G C T G C T G
 751 A T T T T C C T G A A T A C C G G C G T A T C G G C C C T C A T C A G C G A A A A A C T C G T A A G
 801 T G C G G A C G A A A C C T G G G T T C A G A C G G C A A A A A T A A T C G G T T C G A C A C C G A
 851 T C G C C C T T C T G A T T T C C G T A T T G G T C G C A C T G T T G T C T T G G A C G C A A A
 901 C G C G G C G A A A G C G G C A G C G C G T T G G A A A A A A C C G T G G A C G G C G C A C T C G C
 951 C C C C G T C T G T T C C G T G A T T C T G A T T A C C G G C G C G G G C G G T A T G T C G G C G
 1001 G C G T T T T G C G C G C T T C C G G C A T C G G C A A G G C A C T C G C C G A C A G C A T G G C G
 1051 G A T T T G G G C A T T C C C G T C C T T T T G G G C T G T T T C C T T G T C G C C T T G G C A C T
 1101 G C G T A T C G C G C A A G G T T C G G C A A C C G T C G C C C T G A C C A C C G C C G C C G C G
 1151 T G A T G G C T C C T G C C G T T G C C G C C C G G C T T A C C G A C T G C A G C T C G C C
 1201 T G T A T C G T A T T G G C A A C G G C G G C A G G T T C G G T C G G T T G C A G C C A C T T C A A
 1251 C G A C T C C G G C T T C T G G C T G G T C G G C C G C C T T T G G A C A T G A C G T A C C G A
 1301 C C A C G C T G A A A A C C T G G A C G G T C A A C C A A C C C T C A T C G C A C T C A T C G G C
 1351 T T T G C C T T G T C C G C A C T G C T G T T C G C C A T C G T C T G A

This encodes a protein having amino acid sequence <SEQ ID 588>:

20 1 M D G W T Q T L S A Q T L L G I S A A A I I L I L I L I V K F R I H A L L T L V I V S L L T A L A T
 51 G L P T G S I V N D V L V K N F G G T L G G V A L L V G L G A M L G R L V E T S G G A Q S L A D A L
 101 I R M F G E K R A P F A L G V A S L I F G F P I F F D A G L I V M L P I V F A T A R R M K Q D V L P
 151 F A L A S I G A F S V M H V F L P P H P G P I A A S E F Y G A N I G Q V L I L G L P T A F I T W Y F
 201 S G Y M L G K V L G R T I H V P V P E L L S G G T Q D N D L P K E P A K A G T V V A I M L I P M L L
 251 I F L N T G V S A L I S E K L V S A D E T W V Q T A K I I G S T P I A L L I S V L V A L F V L G R K
 301 R G E S G S A L E K T V D G A L A P V C S V I L I T G A G G M F G G V L R A S G I G K A L A D S M A D L G I P V L L G C
 351 F L V A L A L R I A Q G S A T V A L T T A A L M A P A V A A A G F T D W Q L A C I V L A T A A G S V G C S H F N D S G
 401 C I V L A T A A G S V G C S H F N D S G F W L V G R L L D M D V P T T L K T W T V N Q T L I A L I G F A L S A L L F A I V
 451 F A L S A L L F A I V *

30 ORF140a and ORF140-1 show 99.8% identity over a 461aa overlap:

orfl40-1.pep M D G W T Q T L S A Q T L L G I S A A A I I L I L I L I V K F R I H A L L T L V I V S L L T A L A T G L P T G S I V N D 60
 orfl40a M D G W T Q T L S A Q T L L G I S A A A I I L I L I L I V K F R I H A L L T L V I V S L L T A L A T G L P T G S I V N D 60
 35 orfl40-1.pep I L V K N F G G T L G G V A L L V G L G A M L G R L V E T S G G A Q S L A D A L I R M F G E K R A P F A L G V A S L I F 120
 orfl40a V L V K N F G G T L G G V A L L V G L G A M L G R L V E T S G G A Q S L A D A L I R M F G E K R A P F A L G V A S L I F 120
 40 orfl40-1.pep G F P I F F D A G L I V M L P I V F A T A R R M K Q D V L P F A L A S I G A F S V M H V F L P P H P G P I A A S E F Y G 180
 orfl40a G F P I F F D A G L I V M L P I V F A T A R R M K Q D V L P F A L A S I G A F S V M H V F L P P H P G P I A A S E F Y G 810
 orfl40-1.pep A N I G Q V L I L G L P T A F I T W Y F S G Y M L G K V L G R T I H V P V P E L L S G G T Q D N D L P K E P A K A G T V 240
 45 orfl40a A N I G Q V L I L G L P T A F I T W Y F S G Y M L G K V L G R T I H V P V P E L L S G G T Q D N D L P K E P A K A G T V 240
 orfl40-1.pep V A I M L I P M L L I F L N T G V S A L I S E K L V S A D E T W V Q T A K I I G S T P I A L L I S V L V A L F V L G R K 300
 orfl40a V A I M L I P M L L I F L N T G V S A L I S E K L V S A D E T W V Q T A K I I G S T P I A L L I S V L V A L F V L G R K 300
 50 orfl40-1.pep R G E S G S A L E K T V D G A L A P V C S V I L I T G A G G M F G G V L R A S G I G K A L A D S M A D L G I P V L L G C 360
 orfl40a R G E S G S A L E K T V D G A L A P V C S V I L I T G A G G M F G G V L R A S G I G K A L A D S M A D L G I P V L L G C 360
 55 orfl40-1.pep F L V A L A L R I A Q G S A T V A L T T A A L M A P A V A A A G F T D W Q L A C I V L A T A A G S V G C S H F N D S G 420
 orfl40a F L V A L A L R I A Q G S A T V A L T T A A L M A P A V A A A G F T D W Q L A C I V L A T A A G S V G C S H F N D S G 420
 60 orfl40-1.pep F W L V G R L L D M D V P T T L K T W T V N Q T L I A L I G F A L S A L L F A I V 461
 orfl40a F W L V G R L L D M D V P T T L K T W T V N Q T L I A L I G F A L S A L L F A I V 461

Homology with a predicted ORF from *N.gonorrhoeae*

ORF140 shows 92% identity over a 87aa overlap with a predicted ORF (ORF140ng) from

65 *N.gonorrhoeae*:

[illegible]

The complete length ORF140ng nucleotide sequence <SEQ ID 589> was predicted to encode a protein having amino acid sequence <SEQ ID 590>:

10	1	<u>MDGR</u> <u>TOTLSA</u>	<u>QTL</u> <u>LGISAAA</u>	<u>IIL</u> <u>LILIVK</u>	<u>FRIR</u> <u>ALLTLV</u>	<u>IAS</u> <u>LLTALAT</u>
	51	<u>GLPT</u> <u>GSIVND</u>	<u>VLV</u> <u>KNFGGTL</u>	<u>GGV</u> <u>ALLVGLG</u>	<u>AML</u> <u>GLRVETS</u>	<u>GGA</u> <u>QSLADAL</u>
	101	<u>IRM</u> <u>FGKEKRP</u>	<u>FAP</u> <u>GVASLIF</u>	<u>GFPI</u> <u>FFDAGL</u>	<u>IVML</u> <u>PIVFAT</u>	<u>ARR</u> <u>MQKQDVL</u>
	151	<u>FAL</u> <u>ASVGAFS</u>	<u>VMH</u> <u>VFPPHP</u>	<u>GPIA</u> <u>ASEFYG</u>	<u>ANIG</u> <u>QVLILG</u>	<u>LPTA</u> <u>FTITWYF</u>
15	201	<u>SGY</u> <u>MLGKVLG</u>	<u>RAIH</u> <u>VPVPEL</u>	<u>LSG</u> <u>GTQDSDP</u>	<u>KPEK</u> <u>AKAGTV</u>	<u>VAV</u> <u>MLIPLML</u>
	251	<u>IFL</u> <u>NTGVSAL</u>	<u>ISE</u> <u>KLVSAD</u>	<u>TWV</u> <u>QTAKMIG</u>	<u>STP</u> <u>VALLISV</u>	<u>LA</u> <u>ALLVLRGRK</u>
	301	<u>RGE</u> <u>SGSTLEK</u>	<u>TV</u> <u>DGALAPAC</u>	<u>SV</u> <u>LITGAGG</u>	<u>MFG</u> <u>GVLRASG</u>	<u>IG</u> <u>KALADSMA</u>
	351	<u>DLG</u> <u>IPVLLGC</u>	<u>FL</u> <u>VALALRIA</u>	<u>QGS</u> <u>ATVALTT</u>	<u>AA</u> <u>ALMAPAVA</u>	<u>AAG</u> <u>FTDWQLA</u>
	401	<u>CIV</u> <u>LATAAGS</u>	<u>VG</u> <u>CSHFND</u>	<u>FWL</u> <u>VGR</u> <u>LSDM</u>	<u>DVPT</u> <u>TLKTWT</u>	<u>VN</u> <u>QTLIAFIG</u>
	451	<u>FAL</u> <u>SALLFAI</u>	<u>V*</u>			

20 Further work revealed a variant gonococcal DNA sequence <SEQ ID 591>:

	1	ATGGACGGCC	GGACACAGAC	GCTGTCCGCG	CAAACTTGT	TGGGCATTT
	51	GGCGGCGCA	ATCATCTCA	TTCTGATTT	AATCGTAAA	TTCCGCATCC
	101	GC CGCTGCT	GACATCTGGT	ATCGCCAGCC	TGCTGACGGC	TTTGGAAC
25	151	GGTTTGCCA	CAGGCAGCAT	CGTCAACGAC	GTACTGGTCA	AAAATCTCGG
	201	CGGCACGCT	GCCGCGTGG	CGCTTCTGGT	CGGTCTGGGC	GCAATGCTCG
	251	GACGTTTGGT	AGAAACATCC	GGCGGCGCAC	AGTCGCTGGC	GGACGCGCTG
	301	ATCCGGATGT	TCGGCGAAAA	ACGCGCACCG	TTCCGCTCCG	CGGTTGCCTC
	351	GCTGATTTCT	GGCTTCCGGA	TTTTCTTCGA	TGCGGTACTA	ATCGTCATGC
30	401	TGCCCATCGT	ATTCGCCACC	GCACGGCGCA	TGAAACAGGA	CGTACTGCC
	451	TTCCGCGTTG	CCTCCGTCGG	CGCATTTTCC	GTCATGCACG	TCTTCTCGCC
	501	GCCCCATCCG	GGCCCCGATTG	CCGCTTCCGA	ATTTTACGGC	CGGAACATCG
	551	GCCAGGTTTT	GATTTTGGGT	CTGCCGACCG	CTTCTACAC	TGCTGATTTT
	601	AGCGGCTATA	TGCTCGGCAA	AGTGTGGGG	CGCGCCATCC	ATGTTCCCGT
35	651	TCCCGAACTG	CTCAGCGGCG	GCACGCAAGA	CAGCGACCCG	CCGAAGAAGC
	701	CTGCCAAAGC	AGGAACGGTG	GTCGCCGTCA	TGCTGATTTCC	CATGCTGCTG
	751	ATTTTCTGTA	ATACCGCGT	ATCACGCCCTC	ATCAGCGAAA	AACCTGTAAG
	801	TGCGGACGAA	ACTTGGGTTC	AGACGGCAAA	AATGATCGGT	TCGACACCTG
	851	TCGCCCTTCT	GATTTCCGTA	TTGGCCGCAC	TGTTGGTCTT	GGGACGCAAA
40	901	CGCGCGGAAA	GCGGCGACAC	GTTGGAAAAA	ACCGTGGACG	GCGCACTCGC
	951	CCCCGCTGT	CCCGTGATTC	TGATTACCGG	CGCGGGCGGT	ATGTTCCGGG
	1001	GCGTTTTGCG	CGCTTCCGGC	ATCGGCAAGG	CACTCGCCGA	CAGCATGGCG
	1051	GATTTGGGCA	TTCCCGTCCT	TTGGGGCTGC	TTCTTGTGCG	CCTTGGCACT
	1101	CGGTATCGCG	CAAGGTTCCG	CAACCGTCGC	CTTAGCCACA	GCCGCGCGCG
45	1151	TGATGGCTCC	TGCGGTTGCC	GCCGCGGCT	TTACCGACTG	GACAGCTCGC
	1201	TGTATCGTAT	TGGCAACGGC	GGCAGGTTGC	GTCGGTTGCA	GCCACTTCAA
	1251	CGACTCCGGC	TTCTGGCTGG	TCGGCCGCCT	CTTGATATATG	GACGTACCGA
	1301	CCACGCTGAA	AACCTGGACG	GTCAACCAA	CCCTCATCGC	ATTCATCGGC
	1351	TTTGCCCTGT	CCGCACTGCT	GTTTGCCATC	GTCTGA	

This corresponds to the amino acid sequence <SEQ ID 592; ORF140ng-1>:

50	1	MDGRTQTL	SA QTL	LGISAAA	IILILILIVK	FRIRALLTLV	IASLLTALAT
	51	GLPTGSIVND	VLVKNFGGTL	GGVALLVGLG	AMLGRLVETS	GGAQSLADAL	
	101	IRMFGKEKRAP	FAPGVASLIF	GFPIFFDAGL	IVMLPIVFAT	ARRMKQDVLP	
	151	FALASVGAFS	VMHVFLPPHP	GPAASEFYG	ANIGQVLILG	LPTAFITWYF	
55	201	SGYMLGKVLG	RAIHVPVPEL	LSGGTQDSDP	PKEPAKAGTV	VAVMLIFPMLL	
	251	IFLNTGVSAL	ISEKLVSAD	TWVQTAKMIG	STPVALLISV	LAALLVLGRK	
	301	RGESGSTLEK	TVDGALAPAC	SVLITGAGG	MFGGVLRASG	IGKALADSDMA	
	351	DLGIPVLLGC	FLVALALRIA	QGSATVALTT	AAALMAPAVA	AAGFTDQWLA	
	401	CIVLATAAGS	VGCSEHND	SG FWLVGRL	LDM DVPTTLK	TWT VNTQT	LIAFIG
	451	FALSALLFAI	V*				

60 ORF140ng-1 and ORF140-1 show 96.3% identity over 461aa overlap:

orf140ng-1.pep MDGRTQTLSAQTLGLISAAIILILILIVKFRIRALLTLVIASLLTALATGLPTGSIVND

-337-

```

orfl40-1      MDGWTQTLSAQTLGLGISAAAILILILIVKFRIHALLTLVIVSLLTALATGLPTGSIVND
orfl40ng-1.pep VLVKNFGGTLGGVALLVGLGAMLGRVETSGGAQSLADALIRMFGEKRAPFAPGVASLIF
5 orfl40-1      ILVKNFGGTLGGVALLVGLGAMLGRVETSGGAQSLADALIRMFGEKRAPFALGVASLIF
orfl40ng-1.pep GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASVGAFSVMHVFLPPHPGPIAASEFYG
10 orfl40-1      GFPIFFDAGLIVMLPIVFATARRMKQDVLPFALASIGAFSVMHVFLPPHPGPIAASEFYG
orfl40ng-1.pep ANIGQVLILGLPTAFITWYFSGYMLGKVLGRAIHVPVPELLSGGTQSDPPKEPAKAGTV
15 orfl40-1      ANIGQVLILGLPTAFITWYFSGYMLGKVLGRTHVPVPELLSGGTQDNDLPKEPAKAGTV
orfl40ng-1.pep VAVMLIPMLLIFLNTGVSALISEKLVSADETWQTAKMIGSTPVALLISVLAALLVLGRK
20 orfl40-1      VAIMLIPMLLIFLNTGVSALISEKLVSADETWQTAKIIGSTPIALLISVLAALFVLGRK
orfl40ng-1.pep RGESGSTLEKTVDGALAPACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
25 orfl40-1      RGESGSALEKTVDGALAPVCSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGC
orfl40ng-1.pep FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFNDSG
30 orfl40-1      FLVALALRIAQGSATVALTTAAALMAPAVAAAGFTDWQLACIVLATAAGSVGCSHFNDSG
orfl40ng-1.pep FWLVGRLLDMDVPTTLKTWTVNQTLIAFIGFALSALLFAIV
orfl40-1      FWLVGRLLDMDVPTTLKTWTVNQTLIALIGFALSALLFAIV

```

30 Furthermore, ORF140ng-1 is homologous to an *E.coli* protein:

```

gi|882633 (U29579) ORF_o454 [Escherichia coli] >gi|1789097 (AE000358) o454;
This 454 aa ORF is 34% identical (9 gaps) to 444 residues of an approx. 456 aa
protein GNTP BACLI SW: P46832 [Escherichia coli] Length = 454
Score = 210 bits (529), Expect = 1e-53
35 Identities = 130/384 (33%), Positives = 194/384 (49%), Gaps = 19/384 (4%)

Query: 88 ETSGGAQSLADALIRMFGEKRAPFAPGVASLIFGFPIFFDAGLIVMLPIVFATARRMKQD 147
E SGA+SLA+ R G+KR A +A+ G P+FFD G I++ PI++ A+ K
Sbjct: 80 EHSAGAESLANYSRKLGDKRTIAALTAAFFLGIPVFFDVGFIILAPIIYGFAKVAKIS 139

Query: 148 VLPFALASVGAFSVMHVFLPPHPGPIAASEFYGANIGQVLILGLPTAFITWYFSGYMLGK 207
L F L G +HV +PPHPGP+AA+ A+IG + I+G+ + I GY K
Sbjct: 140 PLKFGPLPVAGIMLTVHVAVPPHPGPVAAAGLLHADIGWLTIIIGIAIS-IPVGVVGYFAAK 198

45 Query: 208 VLGRAIHVPVPELL-----SGGTQSDPPKEPAKAGTVVAVMLIPMLLIFLNTGV 257
++ + + E+L G T+ SD P A V +++IP+ +I T
Sbjct: 199 IINKRQYAMSVEVLEQMQLAPASEEGATKLSDKINPPGVA-LVTSLVIPIAIIMAGT-- 255

Query: 258 SALISEKLVSADETWQTAKMIGSTPXXXXXXXXXXXXXXXXXGRKRGESGSTLEKTVDGALA 317
+S L+ + T ++IGS +RG S + AL
50 Sbjct: 256 ---VSATLMPPSHPLLGLTLQLIGSPMVALMIALVLAFWLLALRRGWSLQHTSDIMGSALP 312

Query: 318 PACSVILITGAGGMFGGVLRASGIGKALADSMADLGIPVLLGCFLVALALRIAQGSXXXX 377
A VIL+TGAGG+FG VL SG+GKALA+ + + +P+L F+++LALR +QGS
55 Sbjct: 313 TAAVVILVTGAGGVFGKLVESGVGKALANMLQMDLPLLPAAFIISLALRASQGS--AT 370

Query: 378 XXXXXXXXXXXXXXXGFTDWQLACIVLATAAGSVGCSHFNDSGFWLVGRLLDMDVPTTLK 437
G Q + LA G +G SH NDSGFV+ + L + V LK
60 Sbjct: 371 VAILTTGGLLSEAVMLNPICQVLTVAACFGGLGASHINDSGFWIVTKYLGSLVADGLK 430

Query: 438 TWTVNQTLIAFIGFALSALLFAIV 461
TWTV T++ F GF ++ ++A++
Sbjct: 431 TWTVLTTILGFTGFLITWCWVAVI 454

```

Based on this analysis, including the identification of the presence of a putative leader sequence

65 (double-underlined) and several putative transmembrane domains (single-underlined) in the

gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 71

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 593>:

```

5      1  ..GATTTCGGCA TATCGCCCGT GTATCTTTGG GTTGCCGCCG CGTTCAAACA
      51  TTTGCTGTCT CCGTGGGCTG CCGACTCATA CGATGTCGCA CGCTTTGCAG
    101  GCGTATTTTT TGCCGTTATC GGACTGACTT CTGCGGCTT TGCCGGTTTC
      151  AACTTTTTTG GCAGACACCA CGGGCGCAC. GTCGTCTGA TTCTCATCGG
      201  CTGTATCGGG CTGATTCCAG TTGCCCATTT CCTCAACCCC GCTGCCGCCG
    10  251  CCTTTGCCGC CGCCGGACTG GTGCTGCACG GTTATTCTTT GGCTCGCCGG
      301  CGCGTGATTG CCGCCTCTTT TCTGCTCGGT ACGGGCTGGA CGCTGATGTC
      351  GTTGGCAGCA GCTTATCCGG CAGCATTGTC CCTGATGCTG CCCTTGCCCG
      401  TACTGATGTT TTTCCGTCCG ..

```

This corresponds to the amino acid sequence <SEQ ID 594; ORF141>:

```

15      1  ..DFGISPVYLW VAAAFKLLS PWAADSYDVA RFAGVFFAVI GLTSCGFAGF
      51  NFLGRHHGRX VVLILIGCIG LIPVAHFLNP AAAFAAAGL VLHGYSLARR
     101  RVIAASFLLG TGWTLMSLAA AYPAAFALML PLPVLMEFRP ..

```

Further work revealed the complete nucleotide sequence <SEQ ID 595>:

```

20      1  ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
      51  AAAGCCGTGG CTGCTGCTGT TGATGGCGTT TGCCTGGTGT TGGCCCGGCG
    101  TGTTTCCCA CGATTGTGG AATCCTGACG AACCTGCCGT CTATACCGCC
      151  GTCGAAGCAC TGGCAGGCAG CCCACCCCC TTGGTTGCC ATCTGTTCGG
      201  TCAAACCGAT TTCGCATAC CGCCCGTGA TCTTTGGGT GCCGCCGCT
      251  TCAAACATTT GCTGTGCCG TGGGCTGCC ACTCATACGA TGCCGCACGC
    25  301  TTTGCAGGCG TATTTTTTGC CGTATCGGA CTGACTTCCT GCGGCTTTGC
      351  CGGTTTCAAC TTTTGGGCA GACACCACGG GCGCAGCGTC GTCCTGATTC
      401  TCATCGGCTG TATCGGCTG ATTCCAGTTG CCCATTTCCT CAACCCCGCT
      451  GCCGCCGCTT TTGCCGCCG CGGACTGGTG CTGCACGGTT ATTCTTTGGC
      501  TCGCCGCGCG GTGATTGCC CCTCTTTCT GCTCGGTACG GGCTGGACGC
    30  551  TGATGTCGTT GGCAGCAGCT TATCCGGCAG CATTGCCCCT GATGCTGCCC
      601  TTGCCCGTAC TGATGTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
      651  GACGGCAGTC GCCTCACTTG CCTTGGCCT GCCGCTTATG ACCGTTTACC
      701  CGCTGCTCTT GGCAAAACG CAGCCCGCGC TGTCGCGCA ATGGCTCGAC
      751  TATCACGTTT TCGGTACGTT CGGCGGCGTG CGGCAGGTC AGACGCGATT
    35  801  CAGTTTGTTT TACTATCTGA AAAACCTGCT TTGTTTGCA TTGCCCGCGC
      851  TGCCCGTGGC GGTTTGGACG GTTGGCCGCA CGCGCTGTT TTCGACCGAC
      901  TGGGGGATTT TGGGCGTCGT CTGGATGCTT GCCGTTTGG TGCTGCTTGC
      951  CGTCAATCCG CAGCGTTTTC AGGATAACCT CGTCTGGCTG CTTCCGCGCG
    40  1001  TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGGCG CGGCGCGGCG
     1051  GCGTTTGTC ACTGTTTCG CATTATGGCG TTCGGAAGT GTTCCGTGTT
     1101  CCTGTGGACG GGCTTTTTCG CCATGAATTA CGGCTGGCCC GCCAAGCTTG
     1151  CCGAACGCGC CGCCTATTT AGCCCGTATT ATGTTCTCTG TATCGATCCC
     1201  ATTCCGATGG CGGTTGCCGT ACTGTTTACA CCCTTGTTGG TGTGGGCGAT
     1251  TACCCGGAAT AACATACGCG GCAGGCAGGC GGTTACCAAC TGGGCGGCAG
    45  1301  GCGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG
     1351  GACGCGGCGA AAAGCCACGC GCCGGTCGTC CGGAGTATGG AGGCATCGCT
     1401  TTCCCGGAA TTGAAACGGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA
     1451  TAGGCGGCGG CGACCTGCAC ACGCGGATTG TTTGGACGCA GTACGGCACA
     1501  TTGCCGCACC GCGTCGCGCA TGTACAATGC CGTACCGCA TCGTCTCCT
    50  1551  GCCCCAAAT GCGGATGCGC CGCAAGGCTG GCAGACGGTT TGGCAGGGTG
     1601  CGCGTCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAAATCGGG
     1651  GAAAATATAT AA

```

This corresponds to the amino acid sequence <SEQ ID 596; ORF141-1>:

```

55      1  MLTYTPPDAR PPAKTHEKPW LLLMAFAWL WPGVFSHDLW NPDEPAVYTA
      51  VEALAGSPTP LVAHLFGQTD FGIPPVYLWV AAFKHLSP WAADSYDAAR
    101  FAGVFFAVIG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL IPVAHFLNPA
     151  AAFAAAGLV LHGYSLARRR VIAASFLLG WTLMSLAAA YPAAFALMLP
     201  LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLAKT QPALFAQWLD

```

5

Homology with a predicted ORF from *N.meningitidis* (strain A)

10

15

3

4

6

-340-

1601 CGCGCCCGCG CAACAAAGAC AGTAAGTTCG CACTGATACG GAAAACCGGG
 1651 GAAATATAT TAAAAACAAC AGATTGA

This encodes a protein having amino acid sequence <SEQ ID 598>:

```

1  MLTYTPPDAR PPAKTHEKPW LLLMAFAWL WPGVFSHDLW NPDEPAVYTA
5  51  VEALAGSPTP LVAHLFGQID FGIPPVYLWV AAFAKHLLSP WAADPYDAAR
    101  FAGVFFAVVG LTSCGFAGFN FLGRHHGRSV VLILIGCIGL IPTVHFLNPA
    151  AAFAAAAGLV LHGYSLARRR VIAASFLLGT GWTLMSLAAA YPAAFALMLP
    201  LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLAKT QPALFAQWLD
    251  DHVFGTGGV RHIQTAFSLF YYLKNLLWFA LPALPLAVWT VCRTRLFSTD
10  301  WGILGVVWML AVLVLAVNP QRFQDNLVWL LPPLALFGAA QDLSLRGAA
    351  AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVPDIDP
    401  IPMAVAVLFT PLWLWAIKTR NIRGRQAVTN WAAGVTLTWA LLMTLFLPLW
    451  DAAKSHAPVV RSMEASLSPE LKRELSDGIE CIDIGGDLH TRIVWTQYGT
15  501  LPHRVGDVQC RYRIVRLPON ADAPQGWQTV WQGARPRNKD SKFALIRKTG
    551  ENILKTTD*
  
```

ORF141a and ORF141-1 show 98.2% identity in 553 aa overlap:

```

or f141a . pep  MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
20 or f141-1      MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP

or f141a . pep  LVAHLFGQIDFGIPPVYLWVAAAFKHLLSPWAADPYDAARFAGVFFAVVGLTSCGFAGFN
or f141-1      LVAHLFGQIDFGIPPVYLWVAAAFKHLLSPWAADSYDAARFAGVFFAVVGLTSCGFAGFN

25 or f141a . pep  FLGRHHGRSVVLILIGCIGLIPTVHFLNPAAAFAAAGLVLHGYSLARRRVIAASFLLGT
or f141-1      FLGRHHGRSVVLILIGCIGLIPVAHFLNPAAAFAAAGLVLHGYSLARRRVIAASFLLGT

30 or f141a . pep  GWTLMSLAAAYPAAAFALMLPLPVLMMFFRPWQSRRLMTAVASLAFALPLMTVYPLLAKT
or f141-1      GWTLMSLAAAYPAAAFALMLPLPVLMMFFRPWQSRRLMTAVASLAFALPLMTVYPLLAKT

35 or f141a . pep  QPALFAQWLDLHDHVFGTGGVRHIQTAFSLFYLKNLLWFALPALPLAVWTVCRTRLFSTD
or f141-1      QPALFAQWLDYHVGFTGGVRHVQTAFSLFYLKNLLWFALPALPLAVWTVCRTRLFSTD

40 or f141a . pep  WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQDLSLRGAAAFVNWFGIMA
or f141-1      WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQDLSLRGAAAFVNWFGIMA

45 or f141a . pep  FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMMAVAVLFTPLWLWAIKTR
or f141-1      FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMMAVAVLFTPLWLWAIKTR

50 or f141a . pep  NIRGRQAVTNWAAGVTLTALLMTLFLPWLDAAKSHAPVVRSMELKRELSDGIE
or f141-1      NIRGRQAVTNWAAGVTLTALLMTLFLPWLDAAKSHAPVVRSMELKRELSDGIE

55 or f141a . pep  CIDIGGDLHTRIVWTQYGTLPHRVGDVQCRYRIVRLPONADAPQGWQTVWQGARPRNKD
or f141-1      CIGIGGDLHTRIVWTQYGTLPHRVGDVQCRYRIVLLPONADAPQGWQTVWQGARPRNKD

or f141a . pep  SKFALIRKGTENI
or f141-1      SKFALIRKIGENI
  
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF141 shows 95% identity over a 140aa overlap with a predicted ORF (ORF141ng) from *N.gonorrhoeae*:

```

60 or f141 . pep  DFGISPVYLWVAAAFKHLLSPWAADSYDVA 30
or f141ng      WNPAEPAVYTAVEALAGSPTPLVAHLFGQIDFGIPPVYLWVAAAFKHLLSPWAADPYDAA 126
  
```

orf141.pep RFAGVFFAVIGLTSCGFAGFNLGRHHGRXVVLILIGCIGLIPVAHFLNPAFAAAGL 90
 ||||| 186
 orf141ng RFAGVFFAVIGLTSCGFAGFNLGRHHGRSVLIHIGCIGLIPVAHFFNPAFAAAGL
 5 orf141.pep VLHGYSLARRRVIAASFLLTGTWTLMSLAAAYPAAFALMLPLPVLMMFFRP 140
 ||||| 246
 orf141ng VLHGYSLARRRVIAASFLLTGTWTLMSLAAAYPAAFALMLPLPVLMMFFRPQSRRLMLTA

An ORF141ng nucleotide sequence <SEQ ID 599> was predicted to encode a protein having amino acid sequence <SEQ ID 600>:

10 1 MPSEAVSARP LCEYLLHLAI RPFLTLMLT YTPPDARPPA KTHEKPWLLL
 51 LMAFAWLWPG VFSHDLWNP EAPVYTAVEA LAGSPTPLVA HLFQGTDFGI
 101 PPVYLWVAAA FKHLSPWAA HPYDAARFAG VFFAVIGLTS CGFAGFNFLG
 151 RHHGRSVVLI HIGCIGLIPV AHFNPAFAA FAAAGLVLHG YSLARRRVIA
 201 ASFLLTGTWT LMSLAAAYPA AFALMLPLPV LMFFRPWQSR RLMLTAVASL
 15 251 AFALPLMTVY PLLAKTQPA LFAQWLNHYV FGTGGVVRHI QRAFSLFHYL
 301 KNLLWFAPPG LPLAVWTVCR TRLFSTDWGI LGIVWMLAVL VLLAFNPQRF
 351 QDNLVWLLPP LALFGAAQLD SLRRGAAAFV NWFGIMAFGL FAVFLWTGFF
 401 AMNYGWPAKL AERAAYFSPY YVPDIDPIPM AVAVLFTPLW LWAIRKRNIR
 451 GRQAVTNWAA GVTLTWALLM TLFLPWLDAA KSHAPVVRSM EASFSPELKR
 20 501 ELSDGIECIG IGGGDLHTRI VWTQYGTLP RVGDVRCRYR IVRLPQNADA
 551 PQGWQTVWQG ARPRNKDSKF ALIRKIGENI LKTTD*

Further work revealed the following gonococcal DNA sequence <SEQ ID 601>:

1 ATGCTGACCT ATACCCCGCC CGATGCCCGC CCGCCCGCCA AAACCCACGA
 51 AAAACCGTGG CTGCTGCTGT TGATGGCGTT TGCCTGGCTG TGGCCCGCGG
 25 101 TGTTTTCCCA CGATTTGTGG AATCCTGCCG AACCTGCCGT CTATACCGCC
 151 GTCGAAGCAC TGGCAGGCAG CCCACCCCC TTGGTTGCC ATCTGTTCCG
 201 TCAAACCGAT TTCGGCATA CGCCCGTGTA TCTTTGGGTT GCCCGCGCAT
 251 TCAAACATTT GCTGTCGCCG TGGGCAGCCG ACCCGTATGA TGCCGCACGC
 301 TTTGCAGCGC TATTTTTTGC CGTTATCGGA CTGACTTCTT GCGGCTTTGC
 30 351 CGGTTTCAAC TTTTGGGCA GACACCAGG GCGCAGCGTT GTTTAATCC
 401 ATATCGGCTG TATCGGGCTG ATTCCGGTTG CCCATTCTCT CAATCCgccc
 451 gcccgcgccc tTGCCCGCCG CGGACTGGTG CTGCacgget actcgcgctgGC
 501 ACGCCGGCGC GTGATTgccc cctctTtccT GCTCGGTACG GGTGGACGT
 551 TGATGTCGCT GCGCGCAGCT TATCCGGCGG CGTTTGCCTG GATGCTGCCC
 35 601 CTGCCCGTGC TGATGTTTTT CCGTCCGTGG CAAAGCAGGC GTTTGATGTT
 651 GACGGCAGTC GCCTCGCTTG CCTTTGCCCT GCCGCTTATG ACCGTTTACC
 701 CGCTGCTCtt gGCAAAAACG CAGCCCGCGC TGTTCGCGCA ATGGCTCAAC
 751 TATCACGTTT TCGGTACGTT cggcgGCGTG CGGCACaTTC AGAggGCatT
 801 Cagtttgttt cactatctgA AAaatctgct ttggttcgca ccgccccggc
 40 851 TGCCGCTGGC GGTTCGACG GTTTCGCCA CACGCCTGTT TTCGACCGAC
 901 TGGGGGATT TGGGCATTGT CTGGATGCTT GCCGTTTGG TGCTGCTCGC
 951 CTTTAATCCG CAGCGTTTC AAGACAACCT CGTCTGGCTG CTGCCCGCGC
 1001 TTGCCCTGTT CGGCGCGGCG CAACTGGACA GCCTGAGGCG CGGCGCGGCG
 1051 GCTTTTGTCA ACTGGTTCGG CATTATGGCG TTCGGGCTGT TTGCCGTGTT
 45 1101 CCTGTGGACG GGCTTTTTCG CCATGAATTA CGGCTGGCCC GCCAAGCTTG
 1151 CCGAACGCGC CGCCTACTTC AGCCCGTATT ACCTTCCCGA CATCGATCCC
 1201 ATTCCGATGG CGGTGCGCT ACTGTTTACA CCCTGTGGC TGTGGGCGAT
 1251 TACCCGGAAG AACATACGCG GCAGGCAGGC GGTACCAAC TGGGCGGCAG
 1301 GCGTTACCCT GACCTGGGCT TTGCTGATGA CGCTGTTCTT GCCGTGGCTG
 50 1351 GACGCGGCGA AAAGCCACGC GCCCGTCGTC CGGAGTATGG AGGCATCGTT
 1401 TTCCCGGAA TTAACACGGG AGCTTTCAGA CGGCATCGAG TGTATCGGCA
 1451 TAGGCGGCGG CGACCTGCAC ACGCGGATTG TTTGGACGCA GTACGGCACA
 1501 TTGCGGCACC GCGTCGCGCA TGTCCGTTGC CGCTACCGTA TCGTCCGCT
 1551 GCCCAAAAC GCGGATGCGC CGCAAGGCTG GCAGACGCTG TGGCAGGCTG
 55 1601 CGCGCCCGCG CAACAAGAC AGTAAGTTTG CACTGATACG GAAAATCGGG
 1651 GAAAATATAT TAAAAACAAC AGATTGA

This corresponds to the amino acid sequence <SEQ ID 602; ORF141ng-1>:

1 MLTYTPPDAR PPAKTHEKPW LLLLMFAWL WPGVFSHDLW NPAEPAVYA
 51 VEALAGSPTP LVAHLFGQTD FGIPPVYLWV AAFAKHLSP WAADPYDAAR
 101 FAGVFFAVIG LTSCGFAGFN FLGRHHGRSV VLIHIGCIGL IPVAHFLNPA
 151 AAFAAAAGLV LHGYSLARRR VIAASFLLT GTWTLMSLAAA YPAAFALMLP
 201 LPVLMFFRPW QSRRLMLTAV ASLAFALPLM TVYPLLAKT QPALFAQWLN
 251 YHVFGTFGGV RHIORAFSLF HYLKNLLWFA PPGLPLAVWT VCRTRLFSTD
 301 WGILGIVWML AVLVLAFNP QRFDNLVWL LPPLALFGAA QLDLRRGAA

351 AFVNWFGIMA FGLFAVFLWT GFFAMNYGWP AKLAERAAFY SPYYVPDIDP
 401 IPMAVAVLFT PLWLWAIIRK NIRGRQAVTN WAAGVTLTWA LLMTLFLPWL
 451 DAAKSHAPVV RSMEASFSPK LKRELSDGIE CIGIGGGDLH TRIVWTQYGT
 501 LPHRVGDVRC RYRIVRLPQN ADAPOGWQTV WQGARPRNKD SKFALIRKIG
 551 ENILKTTD*

ORF141ng-1 and ORF141-1 show 97.5% identity in 553 aa overlap:

```

orfl41ng-1.pep MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPAPAVYTAVEALAGSPTP
orfl41-1        MLTYTPPDARPPAKTHEKPWLLLLMAFAWLWPGVFSHDLWNPDEPAVYTAVEALAGSPTP
10
orfl41ng-1.pep LVAHLFGQTDFGIPPVYLWVAAAFKHLLSPWAADPYDAARFAGVFFAVIGLTCGFGAGFN
orfl41-1        LVAHLFGQTDFGIPPVYLWVAAAFKHLLSPWAADSYDAARFAGVFFAVIGLTCGFGAGFN
15
orfl41ng-1.pep FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAFAAAGLVLHGYSLARRRVIAASFLLGT
orfl41-1        FLGRHHGRSVVLIHIGCIGLIPVAHFLNPAAAFAAAGLVLHGYSLARRRVIAASFLLGT
20
orfl41ng-1.pep GWTLMSLAAAYPAAFALMLPLPVLMEFRPWQSRRLMLTAVASLAFALPLMTVYPLLAKT
orfl41-1        GWTLMSLAAAYPAAFALMLPLPVLMEFRPWQSRRLMLTAVASLAFALPLMTVYPLLAKT
25
orfl41ng-1.pep QPALFAQWLNHYHVFETFGGVRHIQRAFSLFHYLKNLLWFAPPGLPLAVWTVCTRRLFSTD
orfl41-1        QPALFAQWLDYHVFETFGGVRHVQTAFSLFYLLKNLLWFALPALPLAVWTVCTRRLFSTD
30
orfl41ng-1.pep WGILGIVWMLAVLVLLAFNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
orfl41-1        WGILGVVWMLAVLVLLAVNPQRFQDNLVWLLPPLALFGAAQLDSLRRGAAAFVNWFGIMA
35
orfl41ng-1.pep FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMNAVAVLFTPLWLWAIIRK
orfl41-1        FGLFAVFLWTGFFAMNYGWPAKLAERAAFYSPYYVPDIDPIPMNAVAVLFTPLWLWAIIRK
40
orfl41ng-1.pep NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAKSHAPVRSMEASFSPKRELSDGIE
orfl41-1        NIRGRQAVTNWAAGVTLTWALLMTLFLPWLDAKSHAPVRSMEASLSPELRELSDGIE
45
orfl41ng-1.pep CIGIGGGDLHTRIVWTQYGTLPHRVGDVRCRYRIVRLPQNADAPQGWQTVWQGARPRNKD
orfl41-1        CIGIGGGDLHTRIVWTQYGTLPHRVGDVQCRYRIVLLPQNADAPQGWQTVWQGARPRNKD
orfl41ng-1.pep SKFALIRKIGENILKTTDX
orfl41-1        SKFALIRKIGENIX

```

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 72

50 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 603>:

```

1  ..CAATCCGCCA AATGGTTATC GGGCCAACT CTAGTCGGCA CAGCAATTGG
51 GATACGCGGG CAGATAAAGC TTGGCGGCAA CCTGCATTAC GATATATTTA
101 CCGGCCGCGC ATTGAAAAG CCCGAATTT TCCAATCAAG GAAATGGGCA
151 AGCGGTTTC AGGTAGGCTA TACGTTTAA

```

55 This corresponds to the amino acid sequence <SEQ ID 604; ORF142>:

```

1  ..QSAKWLSGQT LVGTAIGIRG QIKLGGNLHY DIFTGRALKK PEFFQSRKWA
51 SGFQVGYTF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 605>:

-343-

1 ATGGATAATT CGGGTAGTGA GCGACAGGA AAATACCAAG GAAATATCAC
 51 TTTCTCTGCC GACAATCCTT TGGGACTGAG TGATATGTTC TATGTAAATT
 101 ATGGACGTTC GATTGGCGGT ACGCCCGATG AGGAAAAGTT TGACGGCCAT
 151 CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCCTTT
 5 CCGTAAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG
 251 CAGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAT
 301 ACTGATTTCG GCTTCAACCG CCTGTTGTAT CGTGATGCCA AACGCAAAAC
 351 CTATCTCAGT GTAAAACTGT GGATGAGGGA AACAAAAAGT TACATTGATG
 401 ATGCCGAAC TACTGTACAA CCGCGTAAAA CTGCGGGTTG GTTGGCAGAA
 10 451 CTTTCCCAACA AAGAATATAT CGGTCGCAGT ACGGCAGATT TTAAGTTGAA
 501 ATATAAACGC GGCACCGGCA TGAAGATGC TCTGCGCGCG CCTGAAGAAG
 551 CCTTTGGCGA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT
 601 GTAAATACTC CTTTTCAAAT CGGTAAACAG CTATTTCCTT ATGACACATC
 651 CGTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAACTGG
 15 701 CTATCGGCGG ACACACACCC GTACGTGGCT TCGACGGTGA AATGAGTTTG
 751 TCTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTTGAGCT GGCAATTTAA
 801 ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGTAGGACAT GTTTCAGGAC
 851 AATCCGCCAA ATGGTTATCG GGCCAACTC TAGTCGGCAC AGCAATTGGG
 901 ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTTAC
 20 951 CGGCCGCGCA TTGAAAAAGC CCGAATTTT CCAATCAAGG AAATGGGCAA
 1001 GCGGTTTCA GGTAGGCTAT ACGTTTAA

This corresponds to the amino acid sequence <SEQ ID 606; ORF142-1>:

1 MDNSGSEATG KYQGNITFSA DNPLGLSDMF YVNYGRSIGG TPDEESFDGH
 51 RKEGGSNNYA VHYSAPFGKW TWAFNHNGYR YHQAIVSGLSE VYDYNKGSYN
 101 TDFGFNRLLY RDAKRKTYLG VKLWMRETKS YIDDAELTVQ RRKTAGWLAE
 151 LSHKEYIGRS TADFKLKYKR GTGMKDALRA PEEAFGEGETS RMKIWTASAD
 201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL
 251 SAERGYYWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWSL GQTLVGTAG
 301 IRGQIKLGGN LHYDIFTGRA LKKPEFFQSR KWASGFQVGY TF*

30 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF142 shows 88.1% identity over a 59aa overlap with a predicted ORF (ORF142ng) from *N.gonorrhoeae*:

35 orf142.pep QSAKWLSGQTLVGTAGIRGQIKLGGNLHY 30
 orf142ng RGWYWRNDSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIGIRGQIKLGGNLHY 313
 orf142.pep DIFTGRALKKPEFFQSRKWASGFQVGYTF 59
 40 orf142ng DIFTGRALKKPEYFQTKKWTGFGVGYSF 342

The complete length ORF142ng nucleotide sequence <SEQ ID 607> is:

1 ATGGATAATT CGGGTAGTGA GCGACAGGA AAATACCAAG GAAATATCAC
 51 TTTCTCTGCC GACAATCCTT TTGGACTGAG TGATATGTTC TATGTAAATT
 101 ATGGACGTTC AATTGGCGGT ACGCCCGATG AGGAAAATTT TGACGGCCAT
 151 CGCAAAGAAG GCGGATCAAA CAATTACGCC GTACATTATT CAGCCCCTTT
 201 CCGTAAATGG ACATGGGCAT TCAATCACAA TGGCTACCGT TACCATCAGG
 251 CCGTTTCCGG ATTATCGGAA GTCTATGACT ATAATGGAAA AAGTTACAAC
 301 ACTGATTTCG GCTTCAACCG CCTGTTGTAT CGTGATGCCA AACGCAAAAC
 351 CTATCTCAGT GTAAAACTGT GGACGAGGGA AACAAAAAGT TACATTGATG
 50 401 ATGCCGAAC TACTGTACAA CCGCGTAAAA CCACAGGTTG GTTGGCAGAA
 451 CTTTCCCAACA AAGGATATAT CGGTCGCAGT ACGGCAGATT TTAAGTTGAA
 501 ATATAAACAC GGCACCGGCA TGAAGATGC TCTGCGCGCG CCTGAAGAAG
 551 CCTTTGGCGA AGGCACGTCA CGTATGAAAA TTTGGACGGC ATCGGCTGAT
 601 GTAAATACTC CTTTTCAAAT CGGTAAACAG CTATTTCCTT ATGACACATC
 55 651 CGTTCATGCA CAATGGAACA AAACCCCGCT AACATCGCAA GACAACTGG
 701 CTATCGGCGG ACACACACCC GTACGTGGCT TCGACGGTGA AATGAGTTTG
 751 CCTGCCGAGC GGGGATGGTA TTGGCGCAAC GATTTGAGCT GGCAATTTAA
 801 ACCAGGCCAT CAGCTTTATC TTGGGGCTGA TGTAGGACAT GTTTCAGGAC
 851 AATCCGCCAA ATGGTTATCG GGCCAACTC TAGCCGGCAC AGCAATTGGG
 60 901 ATACGCGGGC AGATAAAGCT TGGCGGCAAC CTGCATTACG ATATATTTAC
 951 CGGCCGTGCA TTGAAAAAGC CCGAATATTT TCAGACGAAG AAATGGGTAA

1001 CGGGGTTTCA GGTGGTTAT TCGTTTGA

This encodes a protein having amino acid sequence <SEQ ID 608>:

1 MDNSGSEATG KYQGNITFSA DNPFGLSDMF YVNYGRSIGG TPDEENFDGH
 51 RKEGGSNNYA VHYSAPFGKW TWAFNHNGYR YHQAVSGLSE VYDNGKSYN
 101 TDFGFNRLLY RDAKRKTYLS VKLWTRETKS YIDDAELTVQ RRKTTGWLAE
 151 LSHKGYIGRS TADFKLKYKH GTGMKDALRA PEEAFGEGETS RMKIWTASAD
 201 VNTPFQIGKQ LFAYDTSVHA QWNKTPLTSQ DKLAIGGHHT VRGFDGEMSL
 251 PAERGWIWRN DLSWQFKPGH QLYLGADVGH VSGQSAKWLS GQTLAGTAIG
 301 IRGQIKLGGN LHYDIFTGRA LKKPEYFQTK KWTGTFQVGY SF*

- 10 The underlined sequence (aromatic-Xaa-aromatic amino acid motif) is usually found at the C-terminal end of outer membrane proteins.

ORF142ng and ORF142-1 show 95.6% identity over 342aa overlap:

orfl42-1.pep MDNSGSEATGKYQGNITFSADNPLGLSDMFYVNYGRSIGGTPDEESFDGHRKEGGSNNYA
 15 orfl42ng-1 MDNSGSEATGKYQGNITFSADNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYA
 orfl42-1.pep VHYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDNGKSYNTDFGFNRLLYRDAKRKTYLG
 20 orfl42ng-1 VHYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDNGKSYNTDFGFNRLLYRDAKRKTYLS
 orfl42-1.pep VKLWMRETKSYIDDAELTVQRRKTAGWLAE LSHKEYIGRSTADFKLKYKRGTGMKDALRA
 orfl42ng-1 VKLWTRETKSYIDDAELTVQRRKTGWLAE LSHKGYIGRSTADFKLKYKHGTGMKDALRA
 25 orfl42-1.pep PEEAFGEGETSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT
 orfl42ng-1 PEEAFGEGETSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHT
 orfl42-1.pep VRGFDGEMSLSAERGWIWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLVGTAG
 30 orfl42ng-1 VRGFDGEMSLPAERGWIWRNDLSWQFKPGHQLYLGADVGHVSGQSAKWLSGQTLAGTAIG
 orfl42-1.pep IRGQIKLGGNLHYDIFTGRALKKPEFFQSRKWASGFQVGYTF
 35 orfl42ng-1 IRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWTGTFQVGYSF

In addition, ORF142ng is homologous to the HecB protein of *E.chrysanthemi*:

gi|1772622 (L39897) HecB [Erwinia chrysanthemi] Length = 558
 Score = 119 bits (295), Expect = 3e-26
 Identities = 88/346 (25%), Positives = 151/346 (43%), Gaps = 22/346 (6%)
 40 Query: 2 DMSGSEATGKYQGNITFSADNPFGLSDMFYVNYGRSIGGTPDEENFDGHRKEGGSNNYAV 61
 DMSG ++TG+ Q N + + DN FGL+D ++++ G S + + D + G
 Sbjct: 230 DMSGQKSTGEEQLNGSLALDNVFLADQWFIAGHS---SRFATSHDAESLQAG----- 280
 45 Query: 62 HYSAPFGKWTWAFNHNGYRYHQAVSGLSEVYDNGKSYNTDFGFNRLLYRDAKRKTYLSV 121
 +S P+G W +N++ RY + G S F +R+++RD KT ++
 Sbjct: 281 -FSMPYGYWNLGYNYSQSRYRNTFINRDFPWHSTGSDTHRFSLSRVVRDGTMTKTAIAG 339
 50 Query: 122 KLWTRETKSYIDDAELTVQRRKTGWLAE LSHKGYIGRSTADFKLKYKHGTGMKDALRAP 181
 R +Y++ + L RK + ++H + A F Y G +
 Sbjct: 340 TFSQRTGNNYLNGLSPSSSRKLSSVSLGVNHSQKLWGGLATFNPTYNRGRVRLGSETDT 399
 Query: 182 EEAFGEGETSRMKIWTASADVNTPFQIGKQLFAYDTSVHAQWNKTPLTSQDKLAIGGHHTV 241
 +++ E + WT SA P Y S++ Q++ L ++L +GG ++
 55 Sbjct: 400 DKSADPRAEFNKWTLASYYHPV---TDSITYLGSLYGQYSARALYGSEQLTLGGESSI 456
 Query: 242 RGFDGEMSLPAERGWIWRNDLSWQFKP----GHQLYLGA-DVGHVSGQSAKWLSGQTLAG 296
 RGF E RG YWRN+L+WQ G+ ++ A D GH+ + +L G
 Sbjct: 457 RGF-REQYTSNGRGAYWRNELNWQAWQLPVLGNVTFMAAVDGGHLYNHKQDNSTAASLWG 515
 60 Query: 297 TAIGIRGQIKLGGNLHYDIFTGRALKKPEYFQTKKWTGTFQVGYSF 342
 A+G+ + L + G + P + Q V G++VG SF

5 Example 73

10

1	ATGCGGACGA	AATGGTCAGC	AGTGAGAAGC	TGCTTACTTG	GgCGGACACC
51	GCCGACATCG	ATACCGCTTT	GAACCTGTTG	TACCGTTTGC	AAAACTGCGA
101	ATTCTCTAT	GGCGATGAAA	ACGGTCATTG	AGACGGCATC	AATTTGwCGG
151	ACGAGCAATT	CCGGTTGTCT	ATGGAACAAT	TGTCGGCGAG	CGGTAAAGCGG
201	TTATTGGTCG	ATCGGAACGG	TCTGTATCTT	GCCAACGCCA	ATTTCCATCA
251	TGAGGCGCGG	CAAGAGTTGG	GGTTGTTGGC	GGCAGAAGTC	GCACAGATGG
301	AAAAGAAATA	CGGGTGTCTG	ATTAAGAACA	GC..	

```

15      1  MRTKWSAVRS CTWADTADID TALNLLYRLQ KLEFLYGDEN GHSDDGINLXD
      51  EQLPLLMEQL  SGSGKALLVD  RNLGLYLANAN  FHHEAAEELG  LLAAEVAQME
     101  KKYRLLIKNN  ..

```

20	1	ATGGAATCAA	CACTTTCACT	ACAAGCAAAT	TTATATCCCC	GCCTGACTCC
	51	TGCCGGTGCA	TTTTATGCCG	TATCCAGCGA	TGCCCCCAGT	GCCGGTAAAA
	101	CTTTGTGTGA	CAGCCTGTGT	AAAGCAGATG	CGGACGAAAT	GGTCAGCAGT
	151	GAGAAGCTGC	TTACTTGGGC	GGACACCGCC	GACATCGATA	CCGCTTTGAA
	201	CCTGTTGTAC	CTTGTGCAAA	AACTCGAATT	CTCTATGGC	GATGAAAACG
25	251	GTCATTCAGA	CGGCATCAAT	TTGTCGGACG	AGCAATTGCC	GTTGCTGATG
	301	GAACAATTGT	CCGGCAGCGG	TAAGGCGTTA	TGTTGTCGAT	GGAACGGTCT
	351	GTATCTTGCC	AACGCCAATT	TCCATCATGA	GGCGGCGGAA	GAGTTGGGGT
	401	TGTTGGCGGC	AGAAGTCGCA	CAGATGGAAA	AGAAATACCG	GCTGCTGATT
	451	AAGAACAACC	TGTATATCAA	CAATAACGCT	TGGGGCGTTT	GCATCCTTTC
30	501	CGGTCAAGC	GAATTGACAT	TTTTCCCATT	GTATATCGGT	TCAACCAAAAT
	551	TTATTTTGGT	TCTCGGCGGC	ATTCCCGATT	TGGGCAAGA	GCAATTTGTT
	601	ACTTTGGTAA	GGATTTTATA	CCGCCGTTAC	AGCAACCGCG	TGTAA

35

1	MESTLSLQAN	LYPRLTPAGA	FYAVSSDAPS	AGKTLHLSLL	KADADEMVSS
51	EKLLTWADTA	DIDTALNLLY	RLQKLEFLYG	DENGHSDGIN	LSDEQLPLLM
101	EQLSGSGKAL	LVDRNGLYLA	NANFHHEAAE	EIGLLAAEVA	QMEKKYRLLI
151	KNNLYINNNA	WGVCDPGSGS	ELT <u>FFPLYIG</u>	STKFILVIGG	IPDLGKEAFV
201	TLVRILYRRY	SNRV*			

Homology with a predicted ORF from *N.meningitidis* (strain A)

```

                                10      20      30
orfl43.pep                    MRTKWSAVRSCTWADTADIDTALNLLYRLQKLEFL
                                | : : ||| ||||| ||||| |||||
45 orfl43a                     GAFYAVSSDXPSAGKTLHSLLKADADEMVSSEKLLTWAXTADIDTALNLLYRLQKLEFL
                                20      30      40      50      60      70
                                40      50      60      70      80      90
orfl43.pep                    YGDENGHS DGINLXDEQLPLLMEQLSGSGKALLVDNRNGLYLANANFHHAAEEELGLLAAE
50                             ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

orf143a      YGDENGHS DGINLSDEQLPL LMEQLSGSGKALLVDRNGLYLANANFHHEAAEELGLLAAE
              80          90          100          110          120          130
              100          110
orf143.pep   VAQMEKKYRLLIKNN
              |||||
orf143a      VAQMEKKYRLXIKNNLYINNNAWGVCDPSPGQSELTFFFPLYIGSTKFILVIGGIPDLGKEA
              140          150          160          170          180          190

```

10	1	ATGGAATCAA	CANTTTCAC	ACAAGCAAAT	TTATATCNCC	GCCTGACTCC
	51	TGCCGGTGCA	TTTTATGCCG	TATCCAGCGA	TGNCCCCAGT	GCCGGTAAAA
	101	CTTTGTTGCA	CAGCCTGTTG	AAAGCGGGATG	CGGACGAAAT	GGTNAGCAGT
	151	GAGAAGCTGC	TTACCTGGGC	GGANACCGCC	GACATCGATA	CCGCTTTGAA
15	201	CCTGTTGTAC	CGTTTGCAAA	AACTCGAATT	CCTCTATGGC	GATGAAACCG
	251	GTCATTGAGA	CGGCATCAAT	TTGTGCGACG	AGCAATTGCC	GTTGCTGATG
	301	GAACAATTGT	CCGGCAGCGG	TAAGCGGTTA	TTGGTGCATG	GGAACGGTCT
	351	GTATCTTGCC	AACGCCAATT	TCCATCATGA	GGCGGCGGAA	GAGTTGGGGT
20	401	TGTTGGCGGC	AGAAGTCGCA	CAGATGGAAA	AGAAATACCG	GCTGCNNATT
	451	AAGAACAACC	TGATATACAA	CAATAACGCT	TGGGGCGGTT	GGCATCCTTC
	501	CGGTGAGAGC	GAAATGACAT	TTTTCCCATT	GTATATCGGT	TCAACCAAAT
	551	TTATTTTGGT	TATCGGGCGC	ATTCCCAGAT	TGGGCAAGA	GGCATTGT
	601	ACTTTGGTAG	GGATNTTATA	CCNCCNGTTA	CAGCAACCGC	GTGTAACCT
	651	TGGGAGAGAG	GANGGGTTAT	GCAGCAATTA	TTGA	

25

1	MESTXSLQAN	LYXRLTPAGA	FYAVSSDXPS	AGKTLHLSL	KADADEMVSS
51	EKLLTWAXTA	DIDTALNLLY	RLQKLEFLYG	DENGHSDGIN	LSDEQLPLLM
101	EQLSGSGKAL	LVDRNGLYLA	NANFHHEAAE	ELGLLAAEVA	QMEKKYRLXI
151	KNNLYINNNA	WGVCDPDSQS	<u>ELTFEFLYIG</u>	<u>STKFIILVIGG</u>	IPDLGKEAFV
201	TLVRXLYXXL	QOPRVKLGRE	XGLCSNY*		

	orf143a.pep	MESTXSLQANLYXRLTPAGAFYAVSSDXPSAGKTLHSLLLKADADEMVSSEKLLTWAXTA
	orf143-1	MESTLSLQANLYPRLTTPAGAFYAVSSDAPSAGKTLHSLLLKADADEMVSSEKLLTWADTA
35	orf143a.pep	DIDTALNLLYRLQKLEFLYGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA
	orf143-1	DIDTALNLLYRLQKLEFLYGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA
40	orf143a.pep	NANFHHEAAEELGLLAAEVAQMEKKYRLXIKNWNLYINNNAWGVCDPSGQSELTFFFPLYIG
	orf143-1	NANFHHEAAEELGLLAAEVAQMEKKYRLLIKWNLYINNNAWGVCDPSGQSELTFFFPLYIG
	orf143a.pep	STKFILVIGGIPDLGKEAFVTLVRXLY
45	orf143-1	STKFILVIGGIPDLGKEAFVTLVRILY

ORF143 shows 95.5% identity over a 110aa overlap with a predicted ORF (ORF143ng) from *N.gonorrhoeae*:

An ORF143ng nucleotide sequence <SEQ ID 615> was predicted to encode a protein having amino acid sequence <SEQ ID 616>:

-347-

1 MRTKWSAVRS CSRADTADID TALNLLYRLQ KLEFLYGDEN GHSDGINLSO
 51 EQLPLMEQL SGSGKALLVD RNLGLYLANAN FHESAEELG LLAAEVAQME
 101 KKYRLLIRNN LYINNNAWGV CDPGQSELT FFPLYIGSTK FILVIAGIPD
 151 LSKGGICYFG KDFIPPLQOP RVKLGTTGIM RQLLISLED LNNTSTDIIA
 201 SAVISTDGLP MATMLPSHLN SDRVGAISAT LLALGSRVQ ELACGELEQV
 251 MIKKGSGYIL LSQAGKDAVL VLVAKETGRL GLILLDAKRA ARHIAEAI*

Further work revealed the following gonococcal DNA sequence <SEQ ID 617>:

1 ATGGAATCAA CACTTTCAC TACAAGCGAAT TTATATCCCT GCCTGACTCC
 51 TGCCGGTGCA TTTTATGCCG TATCCAGCGA TGCCCCCAGT GCCGGTAAAA
 101 CTTTGTTCG CAGCCTGTTG AAAGCGGATG CGGACGAAGT GGTCAGCAGT
 151 GAGAAGCTGC TCGCGGCGGA CACCGCGGAC ATCGATACCG CTTTGAACCT
 201 GTTGTACCGT TTGCAAAAAC TCGAATTCCT CTATGGCGAT GAAAACGGTC
 251 ATTCAGACGG CATCAATTG TCGGACGAGC AATTGCCGTT GCTGATGGAA
 301 CAATTGTCCG GCAGCGGTAA GGCATTATTG GTCGATCGGA ACGGTCTGTA
 351 TCTTGCCAAC GCCAATTTC ATCATGAGTC GCGGGAAGAG TTGGGGTTGT
 401 TGGCGGCAGA AGTCGCACAG ATGGAAAAGA AATACCGGCT GCTGATTAGG
 451 AACAACTGT ATATCAACAA TAACGCTTGG GCGGTTTGGC ATCCTTCGGG
 501 TCAGAGCGAA TTGACATTT TCCCATTTGA TATCGGTTCA ACCAAATTTA
 551 TTTTGGTTAT CGCCGGCATT CCCGATTGA GCAAAGAGGC ATTGTGTACT
 601 TTGGTAAGGA TTTTATACCG CCGTTACAGC AACCGCGTGT AA

This corresponds to the amino acid sequence <SEQ ID 618; ORF143ng-1>:

1 MESTLSLQAN LYPCLTPAGA FYAVSSDAPS AGKTLRLSLL KADAEDEVSS
 51 EKLLAADTAD IDTALNLLYR LQKLEFLYGD ENGHSDGINL SDEQLPLLME
 101 QLSGSGKALL VDRNGLYLAN ANFHESAEEL LGLLAAEVAQ MEKKYRLLIR
 151 NNLYINNNAW GVCDPGQSE LTFFFPLYIGS TKFILVIAGI PDLSKEAFVT
 201 LVRILYRRYS NRV*

ORF143ng-1 and ORF143-1 show 95.8% identity in 214 aa overlap:

orf143ng-1.pep MESTLSLQANLYPCLTPAGAFYAVSSDAPSAGKTLRLSLLKADAEDEVSSSEKLLA-ADTA 59
 10 orf143-1 MESTLSLQANLYPRLTPAGAFYAVSSDAPSAGKTLRLSLLKADAEDEVSSSEKLLTWADTA 60
 orf143ng-1.pep DIDTALNLLYRLQKLEFLYGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA 119
 35 orf143-1 DIDTALNLLYRLQKLEFLYGDENGHS DGINLSDEQLPLLMEQLSGSGKALLVDRNGLYLA 120
 orf143ng-1.pep NANFHESAEELGLLAAEVAQMEKKYRLLIRNNLYINNNAWGVCDPGQSELTFFPLYIG 179
 orf143-1 NANFHESAEELGLLAAEVAQMEKKYRLLIKNNLYINNNAWGVCDPGQSELTFFPLYIG 180
 40 orf143ng-1.pep STKFILVIAGIPDLSKEAFVTLVRILYRRYSNRV 213
 orf143-1 STKFILVIGGIPDLGKEAFVTLVRILYRRYSNRV 214

Based on the presence of the putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 74

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 619>:

1 ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
 51 GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGr
 101 CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCGCTGCTG
 151 ACCGTGATGG TGGCGGTCGC TTCGATTTTC CCCGTGTTTC ACCGCTGGTC
 201 GGATTCGTTT GTCTCCTTCG TCAACCAAAC CATTGTGCCG CA.GGCGCGG
 251 ACATGGTGTG CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
 301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTTCG
 351 GACGATAGAC AATACGTTC ACGCATCTG GCGGGTCAA wTyCCAGCGT
 55 401 CCGTGGATG..

This corresponds to the amino acid sequence <SEQ ID 620; ORF144>:

```

1  MTFLLRLOGL ADNKICAFW FVRRFDEER VPQXAASMTF TLLALVPVL
51  TVMVAVASIF PVFDRWSDSF VSFVNQITVP XGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSLMLIRTID NTFNRIWRVX XQRPWM...

```

5 Further work revealed the complete nucleotide sequence <SEQ ID 621>:

```

1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
51  GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG
101 CGGCGGCAAG CATGACGTTT ACGACGCTGC TGGCACTCGT CCCCCTGCTG
151 ACCGTGATGG TGGCGGTGCG TTCGATTTTC CCCGTGTTCC ACCGCTGGTC
201 GGATTCGTTT GTCTCCTTCG TCAACCAAAC CATTGTGCCC CAGGGCGCGG
251 ACATGGTGTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCTGA TGCTGATTCG
351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCAGCGCTC
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCGG
15 451 CTGCTTTTGG GCGTGGGCAT TTCCTTTATG GTCGGCTCGG TACAGGATGC
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGGCGG
551 CGACGCTGAC CTTTCATGAC CTTTGTGCTG GGGGGCTGTA CCGCTTCGTG
601 CCAAACCGCT TCGTTCCCGC GCGGCAGGCG TTTGTGCGGG CTTTGGCAAC
651 AGCGTTTGT CTGAAACCG CCGCTCCCT CTTCACTTGG TATATGGGCA
20 701 ATTCGACGG CTACCGCTCG ATTTACGGCG CGTTTGCCGC CGTGCCGTTT
751 TTTCTGTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG CCGGCGCGGT
801 GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGGGCT
851 TCGACTCGCG CGGACGGTTT GACGACGTGT TGAATACCTT GCTGCTTCTG
901 GATGCGGCGC AAAAAGAAG CAAAGCCTTG CCTGTTTCAAG AGTTCAGACG
25 951 GCATATCAAT ATGGGCTACG ACAGATTGGG CGAGCTTTTG GAAAAGCTGG
1001 CGCGGCACGG CTACATCTAT TCCGCGAGAC AGGGTTGGGT GTTGAAAACG
1051 GGGGCGGATT CGATTGAGTT GAACGAATC TCAAGCTCT TCGTTTACCG
1101 TCGGTTGCCT GTGAAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGACACCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT
30 1201 CAGGCGAAAA AACGGCAGTA G

```

This corresponds to the amino acid sequence <SEQ ID 622; ORF144-1>:

```

1  MTFLLRLOGL ADNKICAFW FVRRFDEER VPQXAASMTF TLLALVPVL
51  TVMVAVASIF PVFDRWSDSF VSFVNQITVP QGADMVFDYI NAFREQANRL
35 101 TAIGSVMLVV TSLMLIRTID NTFNRIWRVN SQRPMWQFL VYWALLTFGP
151 LSLGVGISFM VGSVQDAALA SGAPQWSGAL RTAATLTFTM LLLWGLYRFV
201 PNRFPARQA FVGALATAFC LETARSLFTW YMGNFDGYRS IYGAFAAVPF
251 FLLWNLNLLT LVLGGAVLTS SLSYWGGEAF RRGFDSRGRF DDVLKILLLL
301 DAAQKEGKAL PVQEFRRHIN MGYDELGELL EKLAHGYIY SGRQGWVLT
40 351 GADSIENEL FKLFFVYRPLP VERDHVNQAV DAVMTFCLQT LNMTLAEFDA
401 QAKKRQ*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF144 shows 96.3% identity over a 136aa overlap with an ORF (ORF144a) from strain A of *N.*

meningitidis:

```

45      10      20      30      40      50      60
orf144.pep  MTFLLRLOGLADNKICAFWFVRRFDEERVVPQXAASMTFTLLALVPVLTVMVAVASIF
          |||
orf144a     MTFLLRLOGLADNKICAFWFVRRFDEERVVPQXAASMTFTLLALVPVLTVMVAVASIF
          |||
50      70      80      90     100     110     120
orf144.pep  PVFDRWSDSFVSFVNQITVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSMLIRTID
          |||
orf144a     PVFDRWSDSFVSFVNQITVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSXMLIRTID
          |||
55      70      80      90     100     110     120
          130
orf144.pep  NTFNRIWRVXXQRPWM
          |||
60  orf144a  NTFNRIWRVNSQRPMWQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL

```

130 140 150 160 170 180

The complete length ORF144a nucleotide sequence <SEQ ID 623> is:

```

1  ATGACCTTTT TACAACGTTT GCAAGGTTTG GCAGACAATA AAATCTGTGC
5  51  GTTTGCATGG TTCGTCGTCC GCCGCTTTGA TGAAGAACGC GTACCGCAGG
101 CGGCGGCAAG CATGACGTTT ACGACACTGC TGGCACTCGT CCCCGTGGCTG
151 ACCGTGATGG TGGCGGTCGC TTCGATTTTC CCCGTGTCG ACCGNTGGTC
201 GGATTCGTTT GTCTCCTTCG TCAACCAAAC CATTGTGCCG CAGGGCGCGG
251 ACATGGTNTT CGACTATATC AATGCGTTCC GCGAGCAGGC GAACCGGCTG
301 ACGGCAATCG GCAGCGTGAT GCTGGTCGTT ACCTCGCNGA TGCTGATTCTG
10 351 GACGATAGAC AATACGTTCA ACCGCATCTG GCGGGTCAAT TCCCAGCGTC
401 CGTGGATGAT GCAGTTTCTC GTCTATTGGG CTTTACTGAC GTTCGGGCCG
451 CTGTCTTTGG GCGTGGGCAT TTCCTTTATN GTCGGCTCGG TACAGGATGC
501 CGCGCTTGCC TCAGGTGCGC CGCAGTGGTC GGGCGCGTTG CGAACGCGCG
551 CGACGCTGAN CTTTCATGACG CTTTGTCTGT GGGGGCTGTA CCGCTNCGTG
15 601 CCAAACGCTC TCCTTCCCGC GCGGCANGCG TTTGTGCGGG CTTTGGCAAC
651 AGCGTTCTGT CTGGAACCGC CGCGTTCCTT CTTTACTTGG TATATGGGCA
701 ATTTTCGACG CTACCGCTCG ATTTACGGNG CGTTTGCCCG CGTGCGCTTT
751 TTTCTGTTGT GGCTGAACCT GTTGTGGACG CTGGTCTTGG GCGGCGCGGT
801 GCTGACTTCT TCACTCTCCT ACTGGCAGGG AGAAGCGTTC CGCAGGNGCT
20 851 TCGACTCGCG CGGACGCTTT GACGACGTGT TGAAATCCT GCTGCTTCTG
901 GATGCGGCGC AAAAAGAAGG CNAAGCCTTG CCTGTTCAGG AGTTCAGACG
951 GCATATCAAT ATGGGCTACG ACGAGTTGGG CGAGCTTTTG GAAAGCTGG
1001 CGCGGCACGG CTACATCTAT TCCGGCAGAC AGGGTGGGT GTTGAAACCG
1051 GGGGCGGATT GATTGAGTT GAACGAACTC TCAAGCTCT TCGTTTACCG
25 1101 TCCGTTGCCT GTGGAAAGGG ATCATGTGAA CCAAGCTGTC GATGCGGTAA
1151 TGATGCCGTG TTTGCAGACT TTGAACATGA CGCTGGCAGA GTTTGACGCT
1201 CAGGCGAAAA AACAGCAGCA ATCTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 624>:

```

1  MTFLQRLQGL ADNKICAFW FVRRFDEER VPQAAASMTF TLLALVPVL
30 51  TVMVAVASIF PVFDRWSDSF VSFVNQITVP QGADMVFDYI NAFREQANRL
101 TAIGSVMLVV TSXMLIRTID NTFNRIWRVN SQRPMWQFL VYWALLTFGP
151 LSLGVGISFX VGSVQDAALA SGAPQWSGAL RTAATLXFMT LLLWGLYRXV
201 PNRFPARXA FVGALATAFC LETARSLFTW YMGNFDGYRS IYGAFAAVPF
25 251 FLLWNLNLT LVLGGAVLTS SLSYWQGEAF RRXFDSRGRF DDVLKILLLL
301 DAAQKEGXAL PVQEFRRHIN MGYDELGELL EKLARHGYIY SGRQGWVLKT
35 351 GADSIENLNL FKLFPYRPLP VERDHVNQAV DAVMMPCLQT LNMTLAEFDA
401 QAKKQQQS*

```

ORF144a and ORF144-1 show 97.8% identity in 406 aa overlap:

```

40 orf144a.pep MTFLQRLQGLADNKICAFWVVRREFDEERVQAAASMTFTLLALVPVLTVMVAVASIF
orf144-1 MTFLQRLQGLADNKICAFWVVRREFDEERVQAAASMTFTLLALVPVLTVMVAVASIF

orf144a.pep PVFDRWSDSFVSFVNQITVPQADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID
45 orf144-1 PVFDRWSDSFVSFVNQITVPQADMVFDYINAFREQANRLTAIGSVMLVVTSLMLIRTID

orf144a.pep NTFNRIWRVNSQRPMWQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL
50 orf144-1 NTFNRIWRVNSQRPMWQFLVYWALLTFGPLSLGVGISFXVGSVQDAALASGAPQWSGAL

orf144a.pep RTAATLXFMTLLWGLYRXVFNRFVPRXAFVGALATAFCLETARSLFTWYMGNFDGYRS
orf144-1 RTAATLTFMTLLWGLYRFVPRXAFVGALATAFCLETARSLFTWYMGNFDGYRS

55 orf144a.pep IYGAFAAVPFFLLWNLNLTSLVGGAVLTSSLSYWQGEAFRRXFDSRGRFDDVLKILLLL
orf144-1 IYGAFAAVPFFLLWNLNLTSLVGGAVLTSSLSYWQGEAFRRXFDSRGRFDDVLKILLLL

orf144a.pep DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGWVLKTGADSIENLNL
60 orf144-1 DAAQKEGXALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGWVLKTGADSIENLNL

orf144a.pep FKLFPYRPLPVERDHVNQAVDAVMMPCLQTLNMTLAEFDAQAKKQQQS 408
65 orf144-1 FKLFPYRPLPVERDHVNQAVDAVMMPCLQTLNMTLAEFDAQAKKQQQS 406

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF144 shows 91.2% identity over a 136aa overlap with a predicted ORF (ORF144ng) from *N.gonorrhoeae*:

5	orf144.pep	MTFLQRLQGLADNKICAFWVVRVFDEERVPOXASMTFTLLALVPVLTVMVAVASIF	60
	orf144ng	MTFLQCWQGSADNKICAFWVIRRFSEERVPOAAASMTFTLLALVPVLTVMVAVASIF	60
10	orf144.pep	PVFDRWSDSEVSVFNQITVPXGADMVFDYINAFREQANRLTAIGSVMLVVTSMLLIRTID	120
	orf144ng	PVFDRWSDSEVSVFNQITVPQGADMVFDYIDAFRDQANRLTAIGSVMLVVTSMLLIRTID	120
	orf144.pep	NTFNRIWRVXXQRPWM	136
15	orf144ng	NAFNRIWRVNTQRPWMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL	180

The complete length ORF144ng nucleotide sequence <SEQ ID 625> is predicted to encode a protein having amino acid sequence <SEQ ID 626>:

	1	MTFLQCWQGS	ADNKICAFW	FVIRRFSEER	VPOAAASMTF	TLLALVPVL
20	51	TVMVAVASIF	PVFDRWSDSF	VSVFNQITVP	QGADMVFDYI	DAFRDQANRL
	101	TAIGSVMLVV	TSLMLIRTID	NAFNRIWRVN	TQRPWMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQWADAL	KTAARLAFMT	LLWGLYRFV
	201	PNRFVPARQA	FVGALITAF	LETARFLFTW	YMGNFDGYRS	IYGAFAAVPF
	251	FLWLNLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILL
25	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLT
	351	GADSIELSEL	FKLFVYRPLP	VERDHVNQAV	DAVMTPLQ	LNMTLAEFDA
	401	QAKKQQQS*				

Further work revealed the following gonococcal DNA sequence <SEQ ID 627>:

	1	ATGACCTTTT	TACAACGTTG	GCAAGGTTTG	GCGGACAATA	AAATCTGTGC
30	51	ATTGTCATGG	TTCGTCATCC	GCCGTTTCAG	TGAAGAGCGC	GTACCGCAGG
	101	CAGCGGCGAG	CATGACGTTT	ACGACACTGC	TGGCACTCGT	CCCCGTACTG
	151	ACCGTAATGG	TCGCGGTGCG	TTCGATTTTC	CCCGTGTTTC	ACCGCTGGTC
	201	GGATTCGTT	GTCTCCTTCG	TCAACCAAAC	CATTGTGCCG	CAGGGCGCGG
	251	ATATGGTGTT	CGACTATATC	GACGCATTCC	GCGATCAGGC	AAACCGGCTG
	301	ACCGCCATCG	GCAGCGTGAT	GCTGGTCGTA	ACCTCGCTGA	TGCTGATTTC
35	351	GACGATAGAC	AATGCGTTCA	ACCGCATCTG	GCGGGTTAAC	ACGCAACGCC
	401	CCTGGATGAT	GCAGTTCTCT	GTTTATTGGG	CGTTGCTGAC	TTTCGGGCGT
	451	TTGTCTTTGG	GTGTGGGCAT	TTCTTTATG	GTCCGGTCCG	TTCAAGATCT
	501	CGTACTCTCC	TCCGGAGCGC	AACAATGGGC	GGACGCGTTG	AAGACGGCGG
	551	CAAGGCTGGC	TTTCATGACG	CTTTTGCTGT	GGGGGCTGTA	CCGCTTCTGT
40	601	CCCAACCGCT	TCGTGCCCGC	CCGGCAGGCG	TTTGTCGGAG	CTTTGATTAC
	651	GGCATTCTGC	CTGGAGACGG	CACGTTTCTT	GTTCACCTGG	TATATGGGCA
	701	ATTTGACGCG	CTACCGCTCG	ATTTACGGCG	CATTTGCCGC	CGTGCCGTTT
	751	TTCTGCTGT	GGTAAACCT	GCTGTGGACG	CTGGTCTTGG	GCGGGGCGGT
	801	GCTGACTTCG	TCGCTGTCTT	ATTGGCAGGG	CGAGGCCTTC	CGCAGGGGAT
45	851	TCGACTCGCG	CGGACGGTTT	GACGACGTGT	TGAAAATCCT	GCTGCTTCTG
	901	GATGCGGCGC	AAAAAGAAGG	CCGAACCCTG	TCCGTTTACG	AGTTCAGACG
	951	GCAATATCAAT	ATGGGTTACG	ATGAATTGGG	CGAGCTTTTG	GAAAAGCTGG
	1001	CGCGGTACGG	CTATATCTAT	TCCGGCAGAC	AGGGCTGGGT	TTTGAAAACG
	1051	GGGGCGGATT	CGATTGAGTT	GAGCGAACTC	TTCAAGCTCT	TCGTGTACCG
50	1101	CCCGTTGCct	gtggaAAGG	ATCATGTGAA	CCAAGCTGtc	gaTGCGGTAA
	1151	TGAcgcgctG	TTTGACAGCT	TTGAACATGA	CGCTGGCGGA	GTTTGACGCT
	1201	CAGgcgAAAA	AACAGCAGCA	GTCTTGA		

This encodes a variant of ORF144ng, having the amino acid sequence <SEQ ID 628; ORF144ng-1>:

55	1	MTFLQWQGL	ADNKICAFW	FVIRRFSEER	VPOAAASMTF	TLLALVPVL
	51	TVMVAVASIF	PVFDRWSDSF	VSVFNQITVP	QGADMVFDYI	DAFRDQANRL
	101	TAIGSVMLVV	TSLMLIRTID	NAFNRIWRVN	TQRPWMQFL	VYWALLTFGP
	151	LSLGVGISFM	VGSVQDSVLS	SGAQWADAL	KTAARLAFMT	LLWGLYRFV
	201	PNRFVPARQA	FVGALITAF	LETARFLFTW	YMGNFDGYRS	IYGAFAAVPF
	251	FLWLNLWT	LVLGGAVLTS	SLSYWQGEAF	RRGFDSRGRF	DDVLKILL
60	301	DAAQKEGRTL	SVQEFRRHIN	MGYDELGELL	EKLARYGYIY	SGRQGWVLT

351 GADSIELSEL FKL FVYRPLP VERDHVNOAV DAVMTPCLQT LNMTLAEFDA
401 QAKKQQQS*

ORF144ng-1 and ORF144-1 show 94.1% identity in 406 aa overlap:

```

5  orf144ng-1.pep MTF LQRWQGLADNKICAFAWFVIRRFSEERVPOAAASMTFTTLLALVPVLTVMVAVASIF
   orf144-1      MTF LQRWQGLADNKICAFAWFVIRRFSEERVPOAAASMTFTTLLALVPVLTVMVAVASIF
10 orf144ng-1.pep PVFDRWSDSFVSFVNQTI VPOGADMVFDYIDAFRDQANRLTAIGSVMLVVTSMLIRTID
   orf144-1      PVFDRWSDSFVSFVNQTI VPOGADMVFDYIDAFRDQANRLTAIGSVMLVVTSMLIRTID
15 orf144ng-1.pep NAFNRIWRVNTORPWWMQFLVYWALLTFGPLSLGVGISFMVGSVQDSVLSSGAQQWADAL
   orf144-1      NTFNRIWRVNSQRPWWMQFLVYWALLTFGPLSLGVGISFMVGSVQDAALASGAPQWSGAL
20 orf144ng-1.pep KTAARLAFMTLLLWGLYRFVFNRFVPARQAFVFGALITAFCLTARFLFTWYMGNFDGYRS
   orf144-1      RTAATLTFTMTLLLWGLYRFVFNRFVPARQAFVFGALITAFCLTARSLFTWYMGNFDGYRS
25 orf144ng-1.pep IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL
   orf144-1      IYGAFAAVPFFLLWLNLLWTLVLGGAVLTSSLSYWQGEAFRRGFDSRGRFDDVLKILLLL
30 orf144ng-1.pep DAAQKEGRTL SVQEFRRHINMGYDELGELLEKLARYGYIYSGRQGVWLKTGADSIELSEL
   orf144-1      DAAQKEGKALPVQEFRRHINMGYDELGELLEKLARHGYIYSGRQGVWLKTGADSIELNEL
   orf144ng-1.pep FKL FVYRPLPVERDHVNOAVDAVMT PCLQTLNMTLAEFDAQAKKQQQS
   orf144-1      FKL FVYRPLPVERDHVNOAVDAVMT PCLQTLNMTLAEFDAQAKKRRQ

```

On this basis of this analysis, including the identification of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

35 Example 75

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 629>:

```

1  ..AGACACGCC GCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
51 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
101 GCACCGATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC
40 151 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TCGCCAAAG
201 CCTGCTTGAA ACACGGGAAC ACGGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 630; ORF146>:

```

1  ..RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTDMRQE ISALVILLQR
51 TRRWLDAHE RQHLRQSLLE TREHG*

```

45 Further work revealed the complete nucleotide sequence <SEQ ID 631>:

```

1  ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA
51 CGAACGCTAC CGTACC GCCCTCATCCA CGCCGTCGGG CTCGGCGGGG
101 CCGTCCTGTT CGCCACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC
50 151 GAGTGGATAG GGATGACCGT CTTCGTCGTC CTCGGCATGC TCCAGTTTCA
201 AGGGGCGATT TACTCCAAGG CGGTGGAACG TATGCTCGGC ACGGTCATCG
251 GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCAC
301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCGG
351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCAGGGCTGA
401 CGATGTGTAT GCTCATCGGC GACAACGCA GCGAATGGCT CGACAGCGGA
55 451 CTCATGCGCG CCATGAACGT CTCATCGGC GCGGCCATCG CCATCGCCGC

```

-352-

501 CGCCAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
 551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
 601 AGGCGCATGA CCCGCGAACG CCTCGAGGAG AACATGGCGA AAATGCGCCA
 5 AATCAACGCA CGCATGGTCA AAAGCCGAG CCATCTCGCC GCCACATCGG
 701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCAC
 751 CGTAAATCG TCAACACCAC CGAGTGCTC CTGACCACCG CCGCCAAGCT
 801 GCAATCTCCC AAATCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT
 851 TCACACTGCT CCAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC
 10 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCC TCGTCATCCT GCTGCAACGC
 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG
 1101 CCTGCTTGAA ACACGGGAAC ACGGCTGA

This corresponds to the amino acid sequence <SEQ ID 632; ORF146-1>:

15 1 MNTSQNRNLV SRWLSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG
 51 EWIGMTVFVV LGMLQFQAI YSKAVERMLG TVIGLGAGLG VLWLNQHYFH
 101 GNLLFYLTVG TASALAGWAA VGKNGYVPEML AGITMCMLIG DNGSEWLD SG
 151 LMRAMNVLIG AAIAIAAAKL LPLKSTLMWR FMLADNLADC SKMIAEISNG
 201 RRMTRERLEE NMAKMRQINA RMVKSRSHLA ATSGESRISP AMMEAMQHAH
 20 RKIVNTTELL LTTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTVALING
 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
 351 TRRWLDAHE RQHLRQSLLE TREHG*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N. meningitidis* (strain A)

25 ORF146 shows 98.6% identity over a 74aa overlap with an ORF (ORF146a) from strain A of *N. meningitidis*:

				10	20	30
	orf146.pep			RHARRIRIDTAINPELEALAEHLHYQWQGF		
30	orf146a	KLNGSEIRLLDRHFTLLQTDLQQTVALINGRHARRIRIDTAINPELEALAEHLHYQWQGF				
		280 290 300 310 320 330				
				40	50	60
	orf146.pep	LWLSTDMRQEISALVILLQTRRKWLDAHERQHLRQSLLETREHG				
35						
	orf146a	LWLSTNMRQEISALVILLQTRRKWLDAHERQHLRQSLLETREHSX				
		340 350 360 370				

The complete length ORF146a nucleotide sequence <SEQ ID 633> is:

40 1 ATGAACACCT CGCAACGCAA CCGCCTCGTC AGCCGCTGGC TCAACTCCTA
 51 CGAACGCTAC CGCTACCGCC GCCTCATCCA CGCCGTCCGG CTGGGCGGGG
 101 CCGTCTGTG TCCGACCGCC TCCGCCCGGC TGCTCCACCT CCAACACGGC
 151 GAGTGGATAG GGATGACCGT CTTCGTCTGC CTGGCATGC TCCAGTTTCA
 201 AGGGGCGATT TACTCCAAGG CCGTGGAACG TATGCTCGGC ACGGTCATCG
 251 GGCTGGGCGC GGGTTTGGGC GTTTTATGGC TGAACCAGCA TTATTTCCAC
 45 301 GGCAACCTCC TCTTCTACCT CACCGTCGGC ACGGCAAGCG CACTGGCCCG
 351 CTGGGCGGCG GTCGGCAAAA ACGGCTACGT CCCTATGCTG GCGGGGCTGA
 401 CGATGTGCAT GCTCATCGGC GACAACGGCA GCGAATGGT CGACACGGC
 451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCGGCCATCG CCATCGCCGC
 50 501 CGCCAACTG CTGCCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
 551 CCGACAACCT GACCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
 601 AGGCGCATGA CCCGCGAACG CCTCGAAGAG AACATGGCGA AAATGCGCCA
 651 AATCAACGCA CGCATGGTCA AAAGCCGAG CCACCTCGCC GCCACATCGG
 701 GCGAAAGCCG CATCAGCCCC GCCATGATGG AAGCCATGCA GCACGCCAC
 751 CGTAAATG TCAACACCAC CGAGTGCTC CTGACCACCG CCGCCAAGCT
 55 801 GCAATCTCCC AAATCAACG GCAGCGAAAT CCGGCTGCTT GACCGCCACT
 851 TCACACTGCT CCAACCGAC CTGCAACAAA CCGTCGCCCT TATCAACGGC
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCC TCGTCATCCT GCTGCAACGC
 60 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TGCGCCAAAG
 1101 CCTGCTTGAA ACACGGGAAC ACAGTTGA

-353-

This encodes a protein having amino acid sequence <SEQ ID 634>:

```

1  MNTSQRNRLV SRWLSYERY RYRRLIHAVR LGGAVLFATA SARLLHLQHG
51 EWIGMTVFVV LGMLQFOGAI YSKAVERMLG TVIGLGAGLG VLWLNQHYFH
101 GNLLFYLTVG TASALAGWAA VGKNGYVPM LAGLTMCLIG DNGSEWFDSS
151 LMRAMNVLI GAAIAIAAKL LPLKSTLMWR FMLADNLTDC SKMIAEISNG
201 RRMTRERLEE NMAKMRQINA RMVKSRSILA ATSGESRISP AMMEAMQHAH
251 RKIVNTTELL LTTAAKLQSP KLNSEIRLL DRHFTLLQTD LQQTVALING
301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
351 TRRKWLDAHE RQHLRQSLLE TREHS*

```

10 ORF146a and ORF146-1 show 99.5% identity in 374 aa overlap:

```

orfl46a.pep  MNTSQRNRLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHG
orfl46-1      MNTSQRNRLVSRWLSYERYRYRRLIHAVRLGGAVLFATASARLLHLQHG
15 orfl46a.pep  LGMLQFOGAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLT
orfl46-1      LGMLQFOGAIYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLT
20 orfl46a.pep  VGKNGYVPM LAGLTMCLIGDNGSEWFDSSGLMRAMNVLI GAAIAIAAKL
orfl46-1      VGKNGYVPM LAGLTMCLIGDNGSEWFDSSGLMRAMNVLI GAAIAIAAKL
25 orfl46a.pep  FMLADNLTDCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSILA
orfl46-1      FMLADNLTDCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSILA
30 orfl46a.pep  AMMEAMQHAHRKIVNTTELLTTAAKLQSPKLNSEIRLLDRHFTLLQTD
orfl46-1      AMMEAMQHAHRKIVNTTELLTTAAKLQSPKLNSEIRLLDRHFTLLQTD
35 orfl46a.pep  RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQ
orfl46-1      RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQ
40 orfl46a.pep  RQHLRQSLLETREHSX
orfl46-1      RQHLRQSLLETREHGX

```

Homology with a predicted ORF from *N.gonorrhoeae*

40 ORF146 shows 97.3% identity over a 75aa overlap with a predicted ORF (ORF146ng) from *N.gonorrhoeae*:

```

orfl46.pep  RHARRIRIDTAINPELEALAEHLHYQWQGF 30
orfl46ng    KLNSEIRLLDRHFTLLQTDLQQTAAALINGRHARRIRIDTAINPELEALAEHLHYQWQGF 364
45 orfl46.pep  LWLSTDMRQEISALVILLQRTTRRKWLDAHERQHLRQSLLETREHG 75
orfl46ng    LWLSTNMRQEISALVIPLQRTTRRKWLDAHERQHLRQSLLETREHG 409

```

An ORF146ng nucleotide sequence <SEQ ID 635> was predicted to encode a protein having amino acid sequence <SEQ ID 636>:

```

1  MSGVRFPSPA PIPSTDPPSG SLCFFTFPLQ TASDMNSSOR KRLSGRWLNS
51 YERYRHRRLI HAVRLGGTVL FATALARLLH LQHGWEIGMT VFVVLGMLQF
101 QGAIYSNAVE RMLGTVIGLG AGLGVLWLNQ HYFHGNLLFY LTIGTASALA
151 GWAAVGKNGY VPM LAGLTMCLIGDNGSEW LQSGLMRAMN VLIGAAIAIA
55 201 AAKLLPLKST LMWRFMLADN LADCSKMAE ISNGRRMTRE RLEQNMVQMR
251 QINARMVKS RSHLAATSGE RISPSMMEAM QHAHRKIVNT TELLTTAAK
301 LQSPKLNSE IRLDRHFTL LQTDLQQTAA LINGRHARRI RIDTAINPEL
351 EALAEHLHYQ WQGF LWLSTN MRQEISALVI PLQRTTRKWL DAHERQHLRQ
401 SLETREHG*

```

60 Further work revealed the following gonococcal DNA sequence <SEQ ID 637>:

-354-

1 ATGAACTCCT CGCAACGCAA ACGCCTTTCC GgccGCTGGC TCAACTCCTA
 51 CGAACGCTac cGCCaccGCC GCCTCATACA TGCCGTGCGG CTCGGCggaa
 101 ccgtCCTGTT CGCCACCGCA CTCGCCGgc tACTCCACCT CCAacacggc
 151 gAATGGATAG GGatgaCCGT CTTCGTCGT CTCGGCATGC TCCAGTTCCA
 5 AGGCgcgatt tActccaacg cgggtgGAacg taTgctcgtt acgggtcatcg
 201 ggctgGGCGC GGGTTTGGgc gTTTATGGC TGAACCAGCA TTAttccac
 251 ggcaacCTcc tcttctacct gaccatcggc acggcaagcg cactggccgg
 301 ctGGGCGGCG GTCGGCAAAA acggctacgt ccctatgctg GCGGGGctgA
 351 CGATGTGCAT gctcatcggc gACAACGGCA GCGAATGGCT CGACAGCGGC
 10 451 CTGATGCGCG CGATGAACGT CCTCATCGGC GCCGCCATCG CCATTGCCGC
 501 CGCCAAACTG CTGCGCTGA AATCCACACT GATGTGGCGT TTCATGCTTG
 551 CCGACAACCT GGCCGACTGC AGCAAAATGA TTGCCGAAAT CAGCAACGGC
 601 AGGCGTATGA CGCGCGAAGC TTTGGAGCAG AATATGGTCA AAATGCGCCA
 651 AATCAACGCA CGCATGGTCA AAAGCCGCGC CCACCTCGCC GCCACATCGG
 15 701 GCGAAAGCCG CATCAGCCCC TCCATGATGG AAGCCATGCA GCACGCCAC
 751 CGCAAAATCG TCAACACCAC CGAGCTGCTC CTGACCACCG CCGCCAAGCT
 801 GCAATCTCCC AACTCAACG GCAGCGAAAT CCGGCTGCTC GACCGCCACT
 851 TCACACTGCT CCAAACCGAC CTGCAACAAA CCGCCGCCCT CATCAACGGC
 901 AGACACGCCC GCCGCATCCG CATCGACACC GCCATCAACC CCGAACTGGA
 20 951 AGCCCTCGCC GAACACCTCC ACTACCAATG GCAGGGCTTC CTCTGGCTCA
 1001 GCACCAATAT GCGTCAGGAA ATTTCCGCCC TCGTCATCCT GCTGCAACGC
 1051 ACCCGCCGCA AATGGCTGGA TGCCACGAA CGCCAACACC TCGCCCAAAG
 1101 CCTGCTTGAA ACACGGGAAC ACGGCTGA

This corresponds to the amino acid sequence <SEQ ID 638; ORF146ng-1>:

25 1 MNSSQRKRLS GRWLSYERY RHRRLHAVR LGGTVLFATA LARLLHLQHG
 51 EWIGMTVFVV LGMLQFQGA IYNAVERMLG TVIGLGAGLG VLWLNQHYFH
 101 GNLLFYLTIG TASALAGWAA VGKNGYVPM LAGLTMCLIG DNGSEWLD SG
 151 LMRAMNVLIG AAIATAAAL LPLKSTLMWR FMLADNLADC SKMIAEISNG
 201 RRMTRERLEQ NMVKMRQINA RMVKSRSILA ATSGESRISP SMMEAMQHAH
 30 251 RKIVNTTELL LTAAKLQSP KLNGSEIRLL DRHFTLLQTD LQOTAALING
 301 RHARRIRIDT AINPELEALA EHLHYQWQGF LWLSTNMRQE ISALVILLQR
 351 TRRWLDAHE RQHLRQSLE TREHG*

ORF146ng-1 and ORF146-1 show 96.5% identity in 375 aa overlap

35 orf146-1.pep MNTSQRNRLVSRWLSYERYRRLIHAVRLGGAVLFATASARLLHLQHGEWIGMTVFVV
 orf146ng-1 MNSSQRKRLSGRWLSYERYRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVV
 orf146-1.pep LGMLQFQGA IYSKAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA
 40 orf146ng-1 LGMLQFQGA IYNAVERMLGTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAA
 orf146-1.pep VGKNGYVPM LAGLTMCLIGDNGSEWLD SGLMRAMNVLIGAAIATAAAL LPLKSTLMWR
 orf146ng-1 VGKNGYVPM LAGLTMCLIGDNGSEWLD SGLMRAMNVLIGAAIATAAAL LPLKSTLMWR
 45 orf146-1.pep FMLADNLADCSKMIAEISNGRRMTRERLEENMAKMRQINARMVKSRSILAATSGESRISP
 orf146ng-1 FMLADNLADCSKMIAEISNGRRMTRERLEQNVMKMRQINARMVKSRSILAATSGESRISP
 50 orf146-1.pep AMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNGSEIRLLDRHFTLLQTDLQOTVALING
 orf146ng-1 SMMEAMQHAHRKIVNTTELLLTAAKLQSPKLNGSEIRLLDRHFTLLQTDLQOTAALING
 55 orf146-1.pep RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQTRRWLDAHE
 orf146ng-1 RHARRIRIDTAINPELEALAEHLHYQWQGF LWLSTNMRQEISALVILLQTRRWLDAHE
 orf146-1.pep RQHLRQSLE TREHGX
 60 orf146ng-1 RQHLRQSLE TREHGX

Furthermore, ORF146ng-1 shows homology with a hypothetical *E.coli* protein:

65 sp|P33011|YEEA ECOLI HYPOTHETICAL 40.0 KD PROTEIN IN COBU-SBMC INTERGENIC REGION
 >gi|1736674|gnl|PID|d1016553 (D90838) ORF_ID:o348#20; similar to [SwissProt
 Accession Number P33011] [Escherichia coli] >gi|1736682|gnl|PID|d1016560 (D90839)
 ORF_ID:o348#20; similar to [SwissProt Accession Number P33011] [Escherichia coli]

-355-

>gi1788318 (AE000292) f352; 100% identical to fragment YEEA_ECOLI SW: P33011 but has 203 additional C-terminal residues [Escherichia coli] Length = 352
Score = 109 bits (271), Expect = 2e-23
Identities = 89/347 (25%), Positives = 150/347 (42%), Gaps = 21/347 (6%)

5 Query: 20 YRHRRLIHAVRLGGTVLFATALARLLHLQHGEWIGMTVFVVLGMLQFQGAIIYSNAVERML 79
YRH R++H R+ L + RL + W +T+ V++G + F G + A ER+
Sbjct: 15 YRHYRIVHGTTRVALAFLTLFLIIRLFTIPESTWPLVTMVVIMGPISFWGNVVPRAFERIG 74

10 Query: 80 GTVIGLGAGLGVLWLNQHYFHGNLLFYLTIGTASALAGWAAVGKNGYVPMLAGLTMCMLI 139
GTV+G GL L L L + A L GW A+GK Y +L G+T+ +++
Sbjct: 75 GTVLGSILGLIALQLE---LISLPLMLVWCAAMFLCGWLALGKKPYQGLLIGVTLAIVV 131

15 Query: 140 GDNSEWLD SGLMRAMNVLIGXXXXXXXXXKLLPLKSTLMWRFMLADNLADCSKMI AEISN 199
G E +D+ L R+ +V++G + P ++ + WR LA +L + +++ +
Sbjct: 132 GSPTGE-IDTALWRS G DVILGSL LAM LFTGIW PQRAFIHWRIQLAKSLTEYNRVYQS AFS 190

20 Query: 200 GRRMTRERLEQNMVKMRQINARMVKSRSHLAATSGESRISPSMMEAMQHAHRKIVNXXXX 259
+ R RLE ++ K+ VK R +A S E+RI S+ E +Q +R +V
Sbjct: 191 PNLLEPRLESHLQKLL---TDAVKMRGLIAPASKETRI PKSIYEGIQITINRNLCMLEL 247

25 Query: 260 XXXXXXXXQSPK---LNGSEIRLLDRHFXXXXXXXXXAALINGRHARRIRIDTAINPEL 316
+ LN ++R D AL G +N +
Sbjct: 248 QINAYWATRP SHFVLLNAQKLR--DTQHM MQILLSLVHALYEGNPQPVFANTEKLND AV 305

Query: 317 EALAHL--HYQWQ-----GFLWLSTNMQRQEISALVILLQRTTRK 354
E L + L H+ + G++WL+ ++ L L+ R RK
Sbjct: 306 EELRQLNNHHD LKVVETPIYGVWLN METAHQLELLSNLICRALRK 352

On the basis of this analysis, including the identification of several transmembrane domains in the
30 gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and
their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 76

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 639>

35 1 ..GCCGAAGACA CGCGCGTTAC CGCACAGCTT TTGAGCGCGT ACGGCATTCA
51 GGGCAAACCTC GTCAGTGTGC GCGAACACAA CGAACGCGCAG ATGGCGGACA
101 AGATTGTCGG CTATCTTTCA GACGGCATGG TTGTGGCACA GGTTCGGAT
151 GCGGTACGC CGCCCGTGTG CGACCCGGGC GCGAACTCG CCCGCCGCGT
201 GCGTGAGGCC GGGTTTAAAG TCGTTCCCGT CGTGGGCGCA AC.GCGGTGA
40 251 TGGCGGCTTT GAGCGTGGCC GGTGTGGAAG GATCCGATTT TTATTTCAAC
301 GGTTTGTAC CGCCGAAATC GGGAGAACGC AGGAACTGT TGCCCAAATG
351 GGTGCGGCGC GCGTTTCCTA TCGTCATGTT TGAAACGCCG CACCGCATCG
401 GTGCAGCGCT TGCCGATATG GCGGAAGTGT TCCCGAAGC CCGATTATG
451 CTGCGCGCGC AAATTACGAA AACGTTTGAA ACGTTCTTAA GCGGCACGGT
501 TGGGGAATT CAGACGGCAT TGTCTGCCGA CGGCGACCAA TCGCGCGGCG
45 551 AGATGGTGT GGTGCTTTAT CCGGCGCAGG ATGAAAACA CGAAGGCTTG
601 TCCGAGTCCG CGCAAAACAT CATGAAAATC CTCACAGCCG AGCTGCCGAC
651 CAAACAGGCG GCGGAGCTTG CTGCCAAAT CACGGGCGAG GGAAAGAAAG
701 CTTGTACGA T..

This corresponds to the amino acid sequence <SEQ ID 640; ORF147>:

50 1 ..AEDTRVTAQL LSAYGIQGL VSVREHNERQ MADKIVGYLS DGMVVAQVSD
51 AGTPAVCDPG AKLARRVREA GFKVVPVUGA XAVMAALSA VEGSDIFYFN
101 GFVPPKSGER RKLFAKWVRA AFPIVMFETP HRIGAAALAD AELFERRILM
151 LAREITKTFE TFLSGTVGEI QTALSADGDQ SRGEMVLVLY PAQDEKHEGL
201 SESAQNIMKI LTAELPTKQA AELAAKITGE GKALYD..

55 Further work revealed the complete nucleotide sequence <SEQ ID 641>:

1 ATGTTTCAGA AACATTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
51 ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC
101 GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG
151 CGCGTTACCG CACAGCTTTT GAGCGGTAC GGCATTGAGG GCAAACCTCGT

-356-

201 CAGTGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTGCGCT
 251 ATCTTTTCTAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG
 301 GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG
 5 351 GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA
 401 GCGTGGCCGG TGTGGAAGGA TCCGATTTTT ATTTCAACGG TTTTGTACCG
 451 CCGAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TCGGGGCGGC
 501 GTTTCCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG
 551 CCGATATGGC GGAACGTGTC CCCGAACGCC GATTAATGCT GCGCGCGGAA
 601 ATTACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
 10 651 GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTGTTG
 701 TGCTTTATCC GCGCGAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
 751 CAAACATCA TGAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC
 801 GGAGCTTGCT GCCAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC
 851 TGGCTCTGTC TTGAAAAAC AAATAG

15 This corresponds to the amino acid sequence <SEQ ID 642; ORF147-1>:

1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
 51 RVTAQLLSAY GIQGLVSVR EHNERQMA DKIVGYLSDGMV VAQVSDAGTP
 101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVEG SDFYFNGFVP
 151 PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRMLLARE
 20 201 ITKTFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA
 251 QNIMKILTAE LPTKQAEALA AKITGEGKKA LYDLALSWKN K*

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical protein ORF286 of *E. coli* (accession number U18997)

ORF147 and *E. coli* ORF286 protein show 36% aa identity in 237aa overlap:

25 Orf147: 1 AEDTRVTAQLLSAYGIQGLVSVREHNERQMA DKIVGYLSDGMVVAQVSDAGTPAVCDPG 60
 AEDTR T LL +GI +L ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG
 Orf286: 43 AEDTRHTGLLLQHFGINARLFALHDHNEQQKAETLLAKLQEGQNIALVSDAGTPLINDPG 102
 30 Orf147: 61 AKLARRVREXXXXXXXXXXXXXXXXXXXXXXXXXEGSDFYFNGFVPPKSGERRKLFKAVVRA 120
 L R RE F + GF+P KS RR
 Orf286: 103 YHLVRTCREAGIRVVPPLPGCAAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAEAE 162
 Orf147: 121 AFIIVMFETPHRIGALADMAELFPERR-LMLAREITKTFETFLSGTVGEIQTALSADGD 179
 ++ +E+ HR+ +L D+ + E R ++LARE+TKT+ET VGE+ + D +
 35 Orf286: 163 PRTLIFYESTHRLDLSLEIDIVAVLGESRYVVLARELTKTWETIHGAPVVGELLAWVKEDEN 222
 Orf147: 180 QSRGEMVLVLYPAQDEKHEGLSESAQNIMKILTAE LPTKQAEALAAKITGEGKKALY 236
 + +GEMVL++ + E L A + +L AELP K+AA LAA+I G K ALY
 40 Orf286: 223 RRGEMVLIV-EGHKAQEEDLPADALRTLALLQAEPLKKAALAAEIHGVKKNALY 278

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF147 shows 96.6% identity over a 237aa overlap with ORF75a from strain A of *N. meningitidis*:

45 orf147.pep 10 20 30
 AEDTRVTAQLLSAYGIQGLVSVREHNERQ
 orf75a TLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQGLVSVREHNERQ
 20 30 40 50 60 70
 50 orf147.pep 40 50 60 70 80 90
 MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXVMAALSVA
 orf75a MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA
 80 90 100 110 120 130
 55 orf147.pep 100 110 120 130 140 150
 GVEGSDFYFNGFVPPKSGERRKLFKAVVRAAFPIVMFETPHRIGALADMAELFPERRLM
 orf75a GVAGSDFYFNGFVPPKSGERRKLFKAVVRAAFPIVMFETPHRIGATLADMAELFPERRLM
 140 150 160 170 180 190
 60 orf147.pep 160 170 180 190 200 210
 LAREITKTFETFLSGTVGEIQTALSADGDQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI

```

|||
orf75a  |||:|||||
LAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEKHEGLSESAQNIMKI
200      210      220      230      240      250

                220      230
orf147.pep  LTAELPTKQAAELAAKITGEGKKALYD
|||
orf75a  LTAELPTKQAAELAAKITGEGKKALYDLALSWNKX
260      270      280      290

```

ORF147 shows 94.1% identity over a 237aa overlap with a predicted ORF (ORF147ng) from *N. gonorrhoeae*:

	orf147.pep	AEDTRVTAQLLSAYGIQKGLVSVREHNERQ	30
15	orf147ng	TLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQKGLVSVREHNERQ	85
	orf147.pep	MADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGAXAVMAALSVA	90
20	orf147ng	MADKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPVVGASAVMAALSVA	145
	orf147.pep	GVEGSDFYFNGFVPPKSGERRKLFQKVVRAAFPIVMFETPHRIGAAALADMAELFPERRLM	150
	orf147ng	GVAESDFYFNGFVPPKSGERRKLFQKVVRAAFVVMFETPHRIGATLADMAELFPERRLM	205
25	orf147.pep	LAREITKTFTFLSGTVGEIQTALSADGDQSRGMVLVLYPAQDEKHEGLSESAQNIMKI	210
	orf147ng	LAREITKTFTFLSGTVGEIQTALAADGNQSRGMVLVLYPAQDEKHEGLSESAQNAMKI	265
30	orf147.pep	LTAELPTKQAAELAAKITGEGKKALYD	237
	orf147ng	LAAELPTKQAAELAAKITGEGKKALYDLALSWKNK	300

An ORF147ng nucleotide sequence <SEQ ID 643> was predicted to encode a protein having amino acid sequence <SEQ ID 644>:

35	1	MSVFQTAFFM	FQKHLQKASD	SVVGGTLYVV	ATPIGNLADI	TLRALAVLQK
	51	ADIIICAEDTR	VTAQLLSAYG	IQGRLVSVRE	HNOROMADKV	IGFLSDGLVV
	101	AQVSDAGTGA	KCDPGAKLAR	RVREAGFKVV	PVVGASAVMA	ALSVAGVAES
	151	DFYFNGFVPP	VSGGERKFLFA	KWVRAAFPVV	MFETPHRIGA	TLADMAELFP
	201	ERRIIMLAREI	TKTFETFLSG	TVGEIQTALA	ADGNQSRGEM	VLVLYPAQDE
40	251	KHEGLSESAQ	NAMKILAAEL	PTKQAAELAA	KITGEGKKAL	YDLALSWKNK
	301	*				

Further work revealed the following gonococcal DNA sequence <SEQ ID 645>:

	1	ATGTTTCAGA	AACACTTGCA	GAAAGCCTCC	GACAGCGTCG	TCGGAGGGAC
45	51	ATTATACGTG	GTTGCCACGC	CCATCGGCAA	TTTGGCAGAC	ATTACCCTGC
	101	GCGCTTTGGC	GGTATTGCAA	AAGGCGGACA	TCATTTGTGC	CGAAGACACG
	151	CGCGTTACTG	CGCAGCTTTT	GAGCGCGTAC	GGCATTACAG	CGAGTTGGT
	201	CAGTGTGCGC	GAACACAACG	AGCGGCAGAT	GGCGGACAAG	GTAATCGGTT
50	251	TCCTTTTCAGA	CGGCCTGGTT	GTGGCGCAGG	TTTCCGATGC	GGGTAGCCCG
	301	GCCGTGTGCG	ACCCGGGCGC	GAAACTCGCC	CGCCGCGTGC	GCGAAGCAGG
	351	GTTCAAAGTC	GTTCCCGTCG	TGGGCGCAAG	CGCGGTAATG	GCGGCGTTGA
	401	GTGTGGCCGG	TGTGGCGGAA	TCCGATTTTT	ATTTCAACGG	TTTTGTACCG
55	451	CCGAAATCGG	GCGAACGTAG	GAAATTGTTT	GCCAAATGGG	TGCGGGCGGG
	501	ATTTCTGTGC	GTCATGTTTG	AAACGCCGCA	CCGAATCGGG	GCAACGCTTG
	551	CCGATATGGC	GGATTTGTTT	CCCGAACGCC	GTCTGATGCT	GGCGCGCGAA
	601	ATCACAAGAA	GGAATTGAAC	GTCCTTAAGC	GCAACGGTTG	GGGAAATTCA
60	651	GACGGCATTG	GCGGCGGACG	GCAACCAATC	GCGCGGCGAG	ATGGTGTGTG
	701	TGCTTTATCC	GGCGCAGGAT	GAAAAACACG	AAGGCTTGTC	CGAGTCTGCG
	751	CAAAATGCCA	TGAAAATCCT	TGCGGCCGAG	CTGCGGACCA	AGCAGGCGGG
	801	GGAGCTTGCG	GCCAAGATTA	CAGGTGAGGG	CAAAAGGCT	TTGTACGATT
	851	TGGCACTGTC	GTGGAAAAAC	AAATGA		

This corresponds to the amino acid sequence <SEQ ID 646; ORF147ng-1>:

```

      1 MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
    51 RVTAQLLSAY GIQGRILSVSR EHNERQMA DK VIGFLSDGLV VAQVSDAGTP
    101 AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVAE SDFYFNGFVP
    151 PKSGERRKLF AKWVRAAFPV VMFETPHRIG ATLDMAELF PERRMLLARE
    201 ITKTETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
    251 QNAMKILAAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*
  
```

ORF147ng shows homology to a hypothetical *E.coli* protein:

```

    sp|P45528|YRAL_ECOLI HYPOTHETICAL 31.3 KD PROTEIN IN AGAI-MTR INTERGENIC REGION
    (F286)
    >gi|606086 (U18997) ORF_f286 [Escherichia coli]
    >gi|1789535 (AE000395) hypothetical 31.3 kD protein in agai-mtr intergenic region
    [Escherichia coli] Length = 286
    Score = 218 bits (550), Expect = 3e-56
    Identities = 128/284 (45%), Positives = 171/284 (60%), Gaps = 4/284 (1%)

    Query: 4   KHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAYGIQ 63
              K Q A +S   G LY+V TPIGNLADIT RAL VLQ D+I AEDTR T LL +GI
    Sbjct: 2   KQHQSADNSQ--GQLYIVPTPIGNLADITQRALEVLAQVDLIAAEDTRHTGLLLQHFGIN 59

    Query: 64   GRLVSVREHNERQMA DKVIGFLSDGLVVAQVSDAGTPAVCDPGAKLARRVREAGFKVVPV 123
              RL ++ +HNE+Q A+ ++ L +G +A VSDAGTP + DPG L R REAG +VVP+
    Sbjct: 60   ARLFALHDHNEQQKAETLLAKLQEQNIALVSDAGTPLINDPGYHLVRTCREAGIRVVPL 119

    Query: 124  VGASAVMAALSVAGVAESDFYFNGFVPPKSGERRKLFKWKVRAAFPVVMFETPHRIGATL 183
              G A + ALS AG+ F + GF+P KS RR ++ +E+ HR+ +L
    Sbjct: 120  PGPCAITALSAAGLPSDRFCYEGFLPAKSKGRRDALKAI EAEPRTLIFYESTHRLDLSL 179

    Query: 184  ADMAELFPERR-LMLAREITKTETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQDEK 242
              D+ + E R ++LARE+TKT+ET VGE+ + D N+ +GEMVL++ +
    Sbjct: 180  EDIVAVLGESRYVVLARELTKTWETIHGAPVGE LLAWVKEDENRRKGEMVLIV-EGHKAQ 238

    Query: 243  HEGLSESAQNAMKILAAELPTKQAAELA AKITGEGKKALYDLAL 286
              E L A + +L AELP K+AA LAA+I G K ALY AL
    Sbjct: 239  EEDLPADALRTLALLQAEPLPKAAALAAEIHGVKNALYKYAL 282
  
```

Based on the computer analysis and the presence of a putative transmembrane domain in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 77

40 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 647>

```

      1 ATGAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
      51 AACCGGTCGC ATCCGCTTCT C.GCTGCTTA CTTAGCCATA TGCCTGTCGT
    101 TCGGCATTCT TCCCCAAGCC TGGGCGGGAC AACTTATTT CGGCATCAAC
    151 TACCAATACT ATCGCGACTT TGCCGAAAAT AAAGGCAAGT TTGCAGTCGG
    201 GCGGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAAT
    251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTC GCGTAACGGC
    301 GTGGCGGCAT TGGTGGGCGT ATCAATATAT TGTGAGCGTG GCACATAACG
    351 GCGGCTATAA CAACGTTGAT TTTGGTGGCG AAGGAak.AA tATCCC.GAT
    401 CAACAACGw TTACTTATAA AATTGTGAAA CGGAATAATT ATAAAGCAGG
    451 GACTAAAGGC CATCCTTATG GCGGCGATTA TCATATGCCG CGTTTGATA
    501 AATWTGTCAAC AGATGCAGAA CCTGTTGAAA TGACCAGTTA TATGGATGGG
    551 CGGAAATATA TCGATCAAAA TAATTACCCT GACCGTGTTT GTATTGGGGC
    601 AGGCAGGCAA TATTGGCGAT CTGATGAAGA TGAGCCCAAT AACCAGGAAA
    651 GTTCATATCA TATTGCAAGT .....
    701 ..... GGCTC ACCAATGTTT ATCTATGATG CCCAAAAGCA
    751 AAAGTGTTTA ATTAATGGGG TATTGCAAAC GGGCAACCCC TATATAGGAA
    801 AAAGCAATGG CTTCCAGCTG GTTCGTAAG ATTGGTTCTA TGATGAAATC
    851 TTTGCTGGAG ATACCCATTC AGTATTCTAC GAACCACTC AAAATGGGAA
    901 ATACTCTTTT AACGACGATA ATAATGGCAC AGGAAAAATC AATGCCAAAC
  
```

951 ATGAACACAA TTCTCTGCCT AATAGATTAA AAACACGAAC CGTTCAATTG
 1001 TTTAATGTTT CTTTATCCGA GACAGCAAGA GAACCTGTTT ATCATGCTGC
 1051 AGGTGGTGTC AACAGTTATC GACCCAGACT GAATAATGGA GAAATATTT
 1101 CCTTTATTGA CGAAGGAAAA GGCGAATTGA TACTTACCAG CAACATCAAT
 1151 CAAGGTGCTG GAGGATTATA TTTCCAAGGA GATTTTACGG TCTCGCCTGA
 1201 AAATAACGAA ACTTGCGCAAG GCGCGGGCGT TCATATCAGT GAAGACAGTA
 1251 CCGTTACTTG GAAAGTAAAC GGCGTGGCAA ACGACCGCCT GTCCAAAATC
 1301 GGCAAAGGCA CGCTG.....
 2101GATAAAG
 2151 TGAAGCTTC ATTGACTAAG ACCGACATCA GCGGCAATGT CGATCTTGCC
 2201 GATCAGCTC ATTTAAATCT CACAGGGCTT GCCCACTCA ACGGCAATCT
 2251 TAGTGCAAT GCGGATACAC GTTATACAGT CAGCCACAAC GCCACCCAAA
 2301 ACGGCAACCK TAGCCTCGTG G.sAATGcCC AAGCAACATT TAATCAAGCC
 2351 ACATTAAACG GCAACACATC GGCTTCgGGC AATGCTTCAT TTAATCTAAG
 2401 CGACCACGCC GTACAAAACG GCAGTCTGAC GCTTTCCGGC AACGCTAAGG
 2451 CAAACGTAAAG CCATTCCGCA CTCAACGGTA ATGTCTCCCT AGCCGATAAG
 2501 GCAGTATTCC ATTTTGAAAG CAGCCGCTTT ACCGGACAAA TCAGCGCGGG
 2551 CAAGGATACG GCATTACACT TAAAAGACAG CGAATGGACG CTGCCGTCAG
 2601 GarCGGAATT AGGCAATTTA AACCTTGACA ACGCCACCAT TACaCTCAAT
 2651 TCCGCCTATC GCCACGATGC GGCAGGGGCG CAAACCGGCA GTGCGACAGA
 2701 TGCGCCGCGC CGCCGTTTCG GCGGTTTCGG CCGTTCCTTA TTATmCGTTA
 2751 CACCGCCAAC TTCGGTAGAA TCCCGTTTCA ACACGCTGAC GGTAACCGGC
 2801 AAATTGAACG GTCAGGGAAC ATTCCGCTTT ATGTCGGAAC TCTTCGGCTA
 2851 CCGCAGCGAC AAATTGAAGC TGGCGGAAAG TTCGGAAGGC ACTTACACCT
 2901 TGGCGGTCAA CAATACCGGC AACGAACCTG CAAGCCTCGA ACAATTGACG
 2951 GTAGTGGAAG GAAAAGACAA CAAACCGCTG TCCGAAAACC TTAATTTAC
 3001 CCTGCAAAAC GAACACGTCG ATGCAGGCGC GTGG.....
 3551TTAGAC CGCGTATTTG CCGAAGACCG
 3601 CCGCAACGCC GTTTGGACAA GCGGCATCCG GGACACCAA CACTACCGTT
 3651 CGCAAGATTT CCGCGCCTAC CGCCAACAAA CCGACCTGCG CCAATCGGT
 3701 ATGCAGAAAA ACCTCGGCAG CCGGCGCGTC GGCATCCTGT TTTCGCACAA
 3751 CCGGACCGAA AACACCTTCG ACGACGGCAT CGGCAACTCG GCACGGCTTG
 3801 CCCACGGCGC CGTTTTCGGG CAATACGGCA TCGACAGGTT CTACATCGGC
 3851 ATCAGgCGCG GCGCGGGGTT TTAGCAGCGG CAGCCTTTcA GACGGCATCG
 3901 GAGsmAAAwT CCGCGCGCGC GTGctGCATT ACGGCATTCA GGCACGAtAC
 3951 CCGCGCGGgtt tCgGCGgATt CGGCATCGAA CCGCACATCG GCGCAACGcg
 4001 ctATTTCGTC CAAAAGCGG ATTACCGCTA CGAAAACGTC AATATCGCCA
 4051 CCCCCGGCCT TGCAATCAAC CGcTACCGCG CCGGCATTaA GGCAGATTAT
 4101 TCATTCAAAC CCGCGCAACA CATTCCATC ACGCCTTATT TGAGCCTGTC
 4151 CTATACCGAT GCCGCTTCGG GCAAAGTCCG AACACGCGTC AATACCGCCG
 4201 TATTGGCTCA GGATTTTCGGC AAAACCGCGA GTGCGGAATG GgGCGTAAAC
 4251 GCCGAAATCA AAGGTTTCAC GCTGTCCCTC CACGCTGCCG CCGCCAAAGG
 4301 CCGCAACTG GAAGCGCAAC ACAGCGCGGG CATCAAATTA GGCTACCGCT
 4351 GGTAA...

This corresponds to the amino acid sequence <SEQ ID 648; ORF1>:

1 MKTTDKRTTE THRKAPKTGR IRFXAAYLAI CLSFGILPQA WAGHTYFGIN
 51 YQYYRDFAE NKGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRNG
 101 VAALVGVQYI VSVAHNGGYN NVDFGAEEXN IXDQXRXYTK IVKRNYYKAG
 151 TKGHYPYGGDY HMPRLHKXVT DAEPVEMTSY MDGRKYIDQN NYPDRVRIGA
 201 GRQYWRSDDED EPNNRESSYH IAS.....GS PMFIYDAQKQ
 251 KWLINGVLQT GNPYIGKSNG FQLVRKDWFY DEIFAGDTHS VFYEPRQNGK
 301 YSFNDNNNGT GKINAKHEHN SLPNRLKTRT VQLFNVSLSE TAREPVYHAA
 351 GGVNSYRPRL NNGENISFID EGKGELILTS NINQGAGGLY FQGDFTVSPE
 401 NNETWQAGAV HISEDSTVTW KVNGVANDRL SKIGKGTLL.....
 701DKVTAS LTKTDISGNV DLADHAHLNL TGLATLNGNL
 751 SANGDTRYTV SHNATQNGNX SLVXNAQATF NQATLNGNTS ASGNASFNLS
 801 DHAVQNGSLT LSGNAKANVS HSALNGNVSL ADKAVFHFES SRFTGQISGG
 851 KDTALHLKDS EWTLPXGKEL GNLNLDNATI TLNSAYRHDA AGAQTGSATD
 901 APRRRSRRSR RSLXVTPPT SVESRENTLT VNGKLNQGT FRFMSELFY
 951 RSDKLKLAES SEGTYTLAVN NTGNEPASLE QLTVEGKDN KPLSENLFNT
 1001 LQNEHVDAGA W.....
 1151LDRVFAEDR
 1201 RNAVWTSGIR DTKHYRSQDF RAYRQQTDLR QIGMQKNLGS GRVGILFSHN
 1251 RTENTFDDGI GNSARLAHGA VFGQYIDRF YIGISAGAGF SSGSLSDGIG
 1301 KXKRRRVLHY GIQARYRAGF GGFIEPHIG ATRYFVQKAD YRYENVNIAT
 1351 PGLAFNRYRA GIKADYSFKP AQHISITPYL SLSYTDAAAG KVRTRVNTAV

1401 LAQDFGKTRS AEWGVNAEIK GFTLSLHAAA AKGPQLEAQH SAGIKLGYRW
 1451 *

Further sequencing analysis revealed the complete nucleotide sequence <SEQ ID 649>:

```

5      1 ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCCGAA
      51 AACCGGCGCG ATCCGCTTCT CGCCTGCTTA CTTAGCCATA TGCCTGTCTGT
101    101 TCGGCATTCT TCCCAAGCC TGGCGGGGAC AACTTATT CGGCATCAAC
      151 TACCAATACT ATCGCGACTT TGCCGAAAT AAAGGCAAGT TTGCAGTCGG
201    201 GCGGAAAGAT ATTGAGGTTT ACAACAAAAA AGGGGAGTTG GTCGGCAAAAT
      251 CAATGACAAA AGCCCCGATG ATTGATTTTT CTGTGGTGTG CCGTAACGGC
10     301 GTGGCGGCAT TGGTGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG
      351 CGCCTATAAC AACGTGTGATT TTGGTGCGGA AGGAAGAAAT CCCGATCAAC
401    401 ATCGTTTAC TTATAAAATT GTGAAACGGA ATATTATATA AGCAGGGACT
      451 AAAGGCCATC CTTATGGCGG CGATTATCAT ATGCCGCGTT TGCATAAATT
501    501 TGTCACAGAT GCAGAACCTG TTGAAATGAC CAGTTATATG GATGGGCGGA
15     551 AATATATCGA TCAAAATAAT TACCCTGACC GTGTTCTGAT TGGGGCAGGC
      601 AGGCAATATT GCGCATCTGA TGAAGATGAG CCCAATAACC GCGAAAGTTC
651    651 ATATCATATT GCAAGTGCCT ATTCTGGCT CGTTGGTGGC AATACCTTTG
      701 CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTTAGGTAG TGA AAAAATT
20     751 AAACATAGCC CATATGTTTT TTTACCAACA GGAGGCTCAT TTGGCGACAG
      801 TGGCTCACCA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA
851    851 ATGGGGTATT GCAAACGGGC AACCCCTATA TAGGAAAAAG CAATGGCTTC
901    901 CAGCTGGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
951    951 CCATTACAGT TTCTACGAAC CACGTCAAAA TGGGAAATAC TCTTTTAACG
1001   1001 ACGATAATAA TGGCACAGGA AAAATCAATG CCAAACATGA ACACAATTCT
25     1051 CTGCCTAATA GATTAAAAAC ACGAACCGTT CAATTGTGTTA ATGTTTCTTT
1101   1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGTGTCAACA
1151   1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCTT TATTGACGAA
1201   1201 GGAAAAGGCG AATTGATACT TACCAGCAAC ATCAATCAAG GTGCTGGAGG
1251   1251 ATTATATTTC CAAGGAGATT TTACGGTCTC GCCTGAAAAA AACGAACTT
30     1301 GGCAAGGCGC GGGCGTTCAT ATCAGTGAAG ACAGTACCGT TACTTGGAAA
1351   1351 GTAAACGGCG TGGCAAACGA CCGCCTGTCC AAAATCGGCA AAGGCACGCT
1401   1401 GCACGTTCAA GCCAAAGGGG AAAACCAAGG CTCGATCAGC GTGGGCGACG
1451   1451 GTACAGTCAT TTTGGATCAG CAGGCAGACG ATAAAGGCAA AAAACAAGCC
35     1501 TTTAGTGAAG TCGGCTTGGT CAGCGGCAGG GGTACCGTGC AACTGAATGC
1551   1551 CGATAATCAG TTCAACCCCG ACAAACTCTA TTTCCGCTTT CGCGGCGGAC
1601   1601 GTTTGGATT AAACGGGCAT TCGCTTTCTG TCCACCGTAT TCAAAATACC
1651   1651 GATGAAGGGG CGATGATTGT CAACCACAAT CAAGACAAAG AATCCACCGT
1701   1701 TACCATTACA GGCAATAAAG ATATTGCTAC AACCGGCAAT AACACACGCT
40     1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT
1801   1801 ACGACCAAAA CGAACGGGCG GCTCAACCTT GTTTACAGC CCGCCGACGA
1851   1851 AGACCGCACC CTGCTGCTTT CCGGCGGAAC AAATTTAAAC GGCAACATCA
1901   1901 CGCAACAAAA CCGCAAACTG TTTTTCAGCG GCAGACCAAC ACCGCACGCC
45     1951 TACAATCATT TAAACGACCA TTGGTCGCAA AAAGAGGGCA TTCCTCGCGG
2001   2001 GGAAATCGTG TGGGACAACG ACTGGATCAA CCGCACATT AAAGCGGAAA
2051   2051 ACTTCCAAAT TAAAGCGGGA CAGGCGGTGG TTTCCGCAA TGTTGCCAAA
2101   2101 GTGAAAGGCG ATTGGCATT GAGCAATCAC GCCCAAGCAG TTTTGGTGT
2151   2151 CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC
2201   2201 TGACAAATTG TGTGAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
50     2251 TTGACTAAGA CCGACATCAG CCGCAATGTC GATCTGCCG ATCAGCTCA
2301   2301 TTTAAATCTC ACAGGGCTTG CCACACTCAA CGGCAATCTT AGTGCAATG
2351   2351 GCGATACAG TTATACAGTC AGCCACAACG CCACCCAAAA CCGCAACCTT
2401   2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAAACGG
2451   2451 CAACACATCG GCTTCGGGCA ATGCTTCATT TAATCTAAGC GACCACGCCG
55     2501 TACAAAACGG CAGTCTGACG CTTTCCGGCA ACGCTAAGG AAACGTAAGC
2551   2551 CATTCCGCAC TCAACGGTAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA
2601   2601 TTTTGAAGC AGCCGCTTTA CCGGACAAAT CAGCGCGGCG AAGGATACGG
2651   2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTCAGG CACGGAAATTA
2701   2701 GGCAATTTAA ACCTTGACAA CGCCACCATT AACTCAATT CCGCTATCG
60     2751 CCACGATGCG GCAGGGGCGC AAACCGGCAG TGCGACAGAT GCGCCGCGCC
2801   2801 GCCGTTGCGC CCGTTCGCGC CGTTCCTAT TATCCGTTAC ACCGCCAACT
2851   2851 TCGGTAGAAT CCCGTTTCAA CACGCTGACG GTAAACGGCA AATTGAACGG
2901   2901 TCAGGGAACA TTCGCTTTA TGTGGAACCT CTTGCGCTAC CGCAGCGACA
2951   2951 AATTGAAGCT GCGGGAAGT TCCGAAGGCA CTTACACCTT GCGGTCAAC
65     3001 AATACCGGCA ACGAACCTGC AAGCCTCGAA CAATTGACGG TAGTGAAGG
3051   3051 AAAAGACAAC AAACCGCTGT CCGAAAACCT TAATTTACCC CTGCAAAACG
3101   3101 AACACGTCGA TGCCGGGCGC TGGCGTTACC AACTCATCCG CAAAGACGGC
3151   3151 GAGTTCGCGC TGCATAATCC GGTCAAAGAA CAAGAGCTTT CCGACAAACT
3201   3201 CCGCAAGGCA GAAGCCAAAA AACAGGCGGA AAAAGACAAC GCGCAAGGCC
70     3251 TTGACGCGCT GATTGCGGCC GGGCGCGATG CCGTCGAAAA GACAGAAAGC
3301   3301 GTTGCCGAAC CCGCCCGGCA GGCAGGCGGG GAAAATGTCG GCATTATGCA

```

-361-

5
10
15
20

```

3351 GCGGAGGAA GAGAAAAAC GGGTGCAGGC GGATAAAGAC ACCGCCTTGG
3401 CGAAACAGCG CGAAGCGGAA ACCCGGCCGG CTACCAACCGC CTCCCCCGCG
3451 GCGCGCCGCG CCGCGCGGGA TTGCGCCGAA CTGCAACCCC AACCGCAGCC
3501 CCAACCGCAG CGCGACCTGA TCAGCCGTTA TGCCAATAGC GGTTCGAGTG
3551 AATTTCCGCG CACGCTCAAC AGCGTTTTCG CCGTACAGGA CGAATTAGAC
3601 CGCGTATTTC CCGAAGACCG CCGCAACGCC GTTTGGACAA GCGGCATCCG
3651 GGACACCAA CACTACCGTT CGCAAGATT CCGCGCCTAC CGCCAACAAA
3701 CCGACCTGCG CCAAATCGGT ATGCAGAAAA ACCTCGGCAG CGGGCGCGTC
3751 GGCATCCTGT TTTGCGACAA CCGGACCGAA AACACCTTCG ACGACGGCAT
3801 CCGCAACTCG GCACGGCTTG CCCACGGCGC CGTTTTCGGG CAATACGGCA
3851 TCGACAGGTT CTACATCGGC ATCAGCGCGG GCGCGGGTT TAGCAGCGGC
3901 AGCCTTTCAG ACGGCATCGG AGGCAAAATC CGCCGCCGCG TGCTGCATTA
3951 CCGCATTTCAG GCACGATACC GCGCCGGTTT CCGCGGATTC GGCATCGAAC
4001 CGCACATCGG CGCAACGCGC TATTTCTGTC AAAAAGCGGA TTACCGCTAC
4051 GAAAACGTCA ATATCGCCAC CCCGGCCCTT GCATTCAACC GCTACCGCGC
4101 GGGCATTAAG GCAGATTATT CATTCAAACC GCGCAACAC ATTTCCATCA
4151 CGCCTTATTT GAGCTGTCC TATACCGATG CCGCTTCGGG CAAAGTCCGA
4201 ACACGCGTCA ATACCGCCGT ATTGGCTCAG GATTTCGGCA AAACCCGCG
4251 TCGGGAATGG GCGTAACG CCGAAATCAA AGGTTTCAG CTGTCCCTCC
4301 ACGCTGCCGC CGCAAAGGC CCGCAACTGG AAGCGCAACA CAGCGCGGGC
4351 ATCAAATTAG GCTACCGCTG GTAA

```

This corresponds to the amino acid sequence <SEQ ID 650; ORF1-1>:

25
30
35
40
45
50

```

1 MKTTDKRTE THRKAPKTGR IRFSPAYLAI CLSFGILPOA WAGHTYFGIN
51 YQYYRDFEEN KGFVAVGAKD IEVYNKKGEL VGKSMKAPM IDFSVVSRRNG
101 VAALVGDQYI VSAHNGGYN NVDFGAEGRN PDQHRFTYKI VKRNNYKAGT
151 KGHYPYGGDYH MPRLHKFVTD AEPVEMTSYM DGRKYIDQNN YPDRVRIGAG
201 RQYWRSEDE PNNRESSYHI ASAYSWLVGG NTFQNGSGG GTVNLGSEKI
251 KHSYPYGLPT GGSFGDSGSP MFIYDAQKQK WLINGVLQGT NPYIGKSNGF
301 QLVKRDWFDY EIFAGDTHSV FYEPRQNGKY SFNDNNGTG KINAKHEHNS
351 LPNRLKTRTV QLFNVSLSSET AREPVYHAAG GVNSYRPRLN NGENISFIDE
401 KGKELILTSN INQAGAGLYF QGDFTVSPEN NETWQAGVH ISEDSTVTWK
451 VNGVANDRLS KIGKGLHVQ AKGENQGSIS VGDGTVIDQ QADDKGGKQA
501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIQNT
551 DEGAMIVNHN QDKESTVTIT GNKDIATTGN NNSLDSKKEI AYNGWFGKEK
601 TTKTNGRLNL VYQPAEDRT LLLSGGTNLN GNITQTNGKL FFSGRPTPHA
651 YNHLNDHWSQ KEGIPRGEIV WDNDWINRTF KAENFQIKGG QAVVSRNVAK
701 VKGDWHLNSH AQAVFGVAPH QSHTICTRSD WTGLTNCVEK TITDDKVIAS
751 LTKTDISGNV DLADHAHLNL TGLATLNGNL SANGDTRYTV SHNATQNGNL
801 SLVGNAQATF NQATLNGNTS ASGNASFNLS DHAVQNGSLT LSGNAKANVS
851 HSALNGNVSL ADKAVFHFES SRFTGQISGG KDTALHLKDS EWTLPSTEL
901 GNLNLDNATI TLNSAYRHDA AGAQTGSATD APRRRSRRSR RSLLSVTPPT
951 SVESRFTLT VNGKLNQGT FRFMSELFY RSDKLKLAES SEGTYTLAVN
1001 NTGNEPASLE QLTVEGKDN KPLSENLFY LQNEHVDAGA WRYQLRKDG
1051 EFRLHNPVKE QELSDKLGA EAKKQAEKDN AQSLDALIAA GRDAVEKTES
1101 VAEPAQAGG ENVGIMQAE EKKRVQADKD TALAKQREAE TRPATTAFFR
1151 ARRARRDLPO LQPPQPPQ RDLISRYANS GLSEFSATLN SVFAVQDEL
1201 RVFAEDRRNA VWTSGIRDIT HYRSQDFRAY RQOTDLRQIG MQKNLGSGRV
1251 GILFSHNRT E NTFDDGIGNS ARLAHGAVFG QYGIDRFYIG ISAGAGFSSG
1301 SLSDGIGGKI RRRVLHYGIQ ARYRAGFGGF GIEPHIGATR YFVQKADYRY
1351 ENVNIATPGL AFNRYRAGIK ADYSFKPAQH ISITPYLSLS YTDASGKVR
1401 TRVNTAVLAQ DFGKTRSAEW GVNAEIKGFT LSLHAAAAGK PQLEAQHSAG
1451 IKLGYRW*

```

Computer analysis of these sequences gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

55 ORF1 shows 57.8% identity over a 1456aa overlap with an ORF (ORF1a) from strain A of *N. meningitidis*:

60

```

          10      20      30      40      50      60
orfl.pep MKTTDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPOAWAGHTYFGINYQYYRDFEEN
          |||
orfla     MKTTDKRTTETHRKAPKTGRIRFSPAYLAIICLSFGILPOAWAGHTYFGINYQYYRDFEEN
          10      20      30      40      50      60

          70      80      90     100     110     120
orfl.pep KGKFAVGAKDIEVYNKKGELVGKSMKAPMIDFSVVSRRNGVAALVGVQYIVSAHNGGYN

```


-362-

```

|||||
orfla      KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVSRNGVAALVGDDQYIVSVAHNGGYN
              70      80      90      100     110     120

5          130      140      150      160      170      180
orfl.pep   NVDFGAEGXNIXDQXRXYKIVKRNNYKAGTKGHPYGGDYHMPRLHKXVTDAEPVEMTSY
|||||
orfla      NVDFGAEGXN-PDQHRFSYQIVKRNNYKPDNS-HPYNGDXHMPRLHKFVTDAEPVEMTSD
              130      140      150      160      170      -

10         190      200      210
orfl.pep   MDGRKYIDQNNYPDRVRIGAGRQYWRSDDEP-----NN-----
|||
orfla      MRGNTYSDKKYPERVIRIGSGHHYWRYDDDKHGDLSYSGAWLIGGNTHMQGWGNGGVXSL
180      190      200      210      220      230

15         220      230      240      250      260
orfl.pep   -----RESSYH-----IA-----SGSPMFIYDAQKQKWLINGVLOTGNPYIGKSNGFQLVRK
|||
orfla      SGDVRHANDYGPMPIAGAAGDSGSPMFIYDKTNNKWLINGVLOTGYPYSGRENGFQLIRK
240      250      260      270      280      290

20         270      280      290      300      310      320
orfl.pep   DWFYDEIFAGDTHSVFYEPONGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRTVQLFNV
|||||
orfla      DWFYDDIYRGDTHTVXFEPNSNGHFSFTSNNGTGTVTETNEKVSNP-KLKVQTVRLFDE
300      310      320      330      340      350

30         330      340      350      360      370      380
orfl.pep   SLSETAREPVYHAAGGVNSYRPLNNGENISFIDEGKGELILTSNINQGAGGLYFQGDFT
|||||
orfla      SLNETDKEPVY-AAGGVNQYRPLNNGENLSFIDYNGKLIILSNINQGAGGLYFEGDFT
360      370      380      390      400      410

35         390      400      410      420      430
orfl.pep   VSPENNETWQAGVHI SEDSTVTWKVNGVANDRLSKIGKGTL-----
|||||
orfla      VSPENNETWQAGVHI SEDSTVTWKVNGVANDRLSKIGKGTLHVQAKGENQGSISVGDGT
420      430      440      450      460      470

40         -----
orfl.pep   -----

45         480      490      500      510      520      530
orfla      VILDQQADDKGGKQAFSEIGLXSGRGTVQLNADNQFNPKLYFGFRGRLDLNGHSLSFH

orfl.pep   -----

50         540      550      560      570      580      590
orfla      RIQNTDEGAMIXXHNATTSTVTITGNESITQPSGKNINRLNYSKEIAYNGWFGEKDTTK

orfl.pep   -----

55         600      610      620      630      640      650
orfla      TNGRLNLVYQPAEDRTXLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLGSGWSKMEG

60         -----
orfla      IPQGEIVWDNDWIXRTFKAENFHIQGGQAVISRNVAKVEGDHLSNHAQAVFGVAPHQSH
660      670      680      690      700      710

65         440      450      460      470      480
orfl.pep   -----XXXXDKVTASLTKTDISGNVDLADHAHLNLTGLATLNGNLSAN
|||
orfla      TICTRSDDWTGLTNCVEXXITDDKVIA SLTKTDXSGXVXLXXXXXXXXLXGXAXLXGNLSAN
720      730      740      750      760      770

70         490      500      510      520      530      540
orfla      GDTRYTVSHNATQNGNXLVXNAQATFNQATLNGNTSASGNASFNLSDHAVQNGSLTSG

```

-363-

```

|||||
orfla  GDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNXSXSGNASFNLSNNAQNGSLTSLD
      780      790      800      810      820      830

5      550      560      570      580      590      600
orfl.ppep NAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGKELGNL
orfla     NAKANVSHSALNGNVSLADKAVFHFENSRTGQLSGSKXTALHLKDSEWTLPSGTELGNL
      840      850      860      870      880      890

10     610      620      630      640      650      660
orfl.ppep NLDNATITLNSAYRHDAAGAQTGSATDAPRRRSRRSRLXVTPPTSVESRFNLTVNG
orfla     NLDNATITLNSAYRHDAAGAQTGXVSDTPRRRSRRS---LLSVTPPTSVESRFNLTVNG
      900      910      920      930      940      950

15     670      680      690      700      710      720
orfl.ppep KLNGQGTFRFMSELFGYRSDKLKLAESSEGTYLAVNNTGNEPASLEQLTVEGKDNKPL
orfla     KLNQGTFRFMSELFGYRSDKLKLAESSEGTYLAVNNTGNEPVSLDQLTVEGKDNKPL
      960      970      980      990     1000     1010

20     730      740      750
orfl.ppep SENLNFTLQNEHVDAGAW-----
orfla     SENLNFTLQNEHVDAGAWRYQLIRKDGFEFLHNPVKEQELSDKLKAEAKKQAEKDNAQS
      1020     1030     1040     1050     1060     1070

30     orfl.ppep -----
orfla     LDALIAAGRDAAEKTESVAEPARXAGGENVGIMQAEEEKKRVQADKDSALAKQREAETRP
      1080     1090     1100     1110     1120     1130

35     orfl.ppep -----LDR
orfla     XTAFPRARXARRDLPPQPQPQPQPQORDLXSRYANGLSEFSATLNSVFAVQDELOR
      1140     1150     1160     1170     1180     1190

40     770      780      790      800      810      820
orfl.ppep VFAEDRRNAVWTSGIRDTKHYRSQDFRAYRQOTDLRQIGMQKNLGSGRVGILFSHNRTEN
orfla     VFAEDRRNAVWTSXIRXTKHYRSQDFRAYRQOTDLRQIGMQKNLGSGRVGILFSHNRTEN
      1200     1210     1220     1230     1240     1250

45     830      840      850      860      870      880
orfl.ppep TFDDGIGNSARLAHGA VFGQYIDRFYIGISAGAGFSSGSLSDGIGXKXRRRVLHYGIAQ
orfla     XFDDGIGNSARLAHGA VFGQYIGRFDIGISTGAGFSSGXLSDGIGGKIRRRVLHYGIAQ
      1260     1270     1280     1290     1300     1310

50     890      900      910      920      930      940
orfl.ppep RYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHI
orfla     RYRAGFGGFGIEPYIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPAQHX
      1320     1330     1340     1350     1360     1370

55     950      960      970      980      990     1000
orfl.ppep SITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKP
orfla     SITPYXLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSXHAAAAGKP
      1380     1390     1400     1410     1420     1430

60     1010     1020
orfl.ppep QLEAQHSAGIKLGYRWX
orfla     QLEAQHSAGIKLGYRWX
      1440     1450

```

70 The complete length ORF1a nucleotide sequence <SEQ ID 651> is:

	1	ATGAAAACAA	CCGACAAACG	GACAACCGAA	ACACACCGCA	AAGCCCCGAA
	51	AACCGGCCGC	ATCCGCTTCT	CGCCTGCTTA	CTTAGCCATA	TGCCTGTCTG
	101	TCGGCATTCT	TCCCAAGCT	TGGCGGGAC	ACACTTATTT	CGGCATCAAC
5	151	TACCAATACT	ATCGCGACTT	TGCCGAAAT	AAAGGCAAGT	TTGCAGTCGG
	201	GGCGAAAGAT	ATTGAGGTNT	ACAACAAAA	AGGGGAGTTG	GTCGGCAAAT
	251	CAATGACAAA	AGCCCCGATG	ATTGATTTTT	CTGTGGTGTG	GCGTAACGGC
	301	GTGGCGGCAT	TGGTGGGCGA	TCAATATATT	GTGAGCGTGG	CACATAACGG
	351	CGGCTATAAC	AACGTTGATT	TTGGTGCGGA	AGGAAGNAAT	CCCGATCAGC
	401	ACCGTTTTTC	TTACCAAAAT	GTGAAAAGAA	ATAATTATAA	GCCTGACAA
10	451	TCACACCCTT	ACAACGGCGA	TTANCATATG	CCGCGTTTGC	ATAAATTTGT
	501	CACAGATGCA	GAACCTGTCT	AAATGACGAG	TGACATGAGG	GGGAATACCT
	551	ATTCCGATAA	AGAAAAATAT	CCCGAGCGTG	TCCGCATCGG	CTCAGGACAC
	601	CACATTTGGC	GTTATGATGA	TGACAAACAC	GGCGATTAT	CCTACTCCGG
	651	CGCATGGTTA	ATTGGCGGCA	ATACACATAT	GCAGGGTTGG	GGAAATAATG
15	701	GCGTANTTAG	TTTGAGCGGC	GATGTGCGCC	ATGCCAACGA	CTATGGCCCT
	751	ATGCCGATTG	CAGGTGCGGC	AGGCGACAGC	GGTTCGCCAA	TGTTTATTTA
	801	TGACAAAACA	AACAATAAAT	GGCTGCTCAA	CGGAGTTTTA	CAAACCGGCT
	851	ACCCTTATTC	CGGCAGGGAA	AACGGTTTCC	AGCTGATACG	CAAAGATTGG
	901	TTCTACGATG	ACATTTACAG	AGGCGATACA	CATACCGTCT	NTTTGAACCC
20	951	GCGCAGTAAC	GGACATTTTT	CCTTTACATC	CAACAACAAC	GGTACGGGTA
	1001	CGGTAACAGA	AACCAACGAA	AAGGNTTCCA	ATCCAAAGCT	TAAAGTACAG
	1051	ACAGTCCGAC	TGTTTGACGA	ATCTTTGAAT	GAAACTGATA	AAGAACCAGT
	1101	TTACGCGGCA	GGGGGTGTTA	ATCAGTACCG	TCCAAGGTTA	AACAACGGTG
	1151	AAAACCTTTC	TTTTATCGAT	TACGGCAACG	GCAAACTCAT	CTTATCAAAC
25	1201	AACATCAACC	AAGGCGCGGG	CGGTTTGTAT	TTTGAAGGTG	ATTTTACGGT
	1251	CTCGCCTGAA	AACAACGAAA	CGTGCCAAGG	CGCGGGCGTT	CATATCAGTG
	1301	AAGACAGTAC	CGTTACTTGG	AAAGTAAACG	GCGTGGCAAA	CGACCGCCTG
	1351	TCCAAAATCG	GCAAGGCAC	GCTGCACGTT	CAAGCCAAAG	GGGAAAACCA
	1401	AGGCTCGATC	AGCGTGGGCG	ACGGTACAGT	CATTTTGGAT	CAGCAGGCAG
30	1451	ACGATAAAGG	CAAAAAACAA	GCCTTTAGTG	AAATCGGCTT	GNTCAGCGGC
	1501	AGGGGTACGG	TGCAACTGAA	TGCCGATAAT	CAGTTCAACC	CCGACAACT
	1551	CTATTTGCGC	TTTCGCGGCG	GACGTTTGGA	TTTAAACGGG	CATTCGCTTT
	1601	CGTTCCACCG	TATTCAAAAT	ACCGATGAAG	GGGCGATGAT	TGNCNATCAT
35	1651	AATGCCACAA	CAACATCCAC	CGTTACCATT	ACAGGGAATG	AAAGTATTAC
	1701	ACAACCGAGT	GGTAAGAATA	TCAATAGACT	TAATTACAGC	AAAGAAATTG
	1751	CCTACAACGG	TTGGTTTGGC	GAGAAAGATA	CGACCAAAAC	GAACGGGCGG
	1801	CTCAACCTTG	TTTACCAGCC	CGCCGCAGAA	GACCGCACCC	NGCTGCTTTC
	1851	CGGCGGAACA	AATTTAAACG	GCAACATCAC	GCAAAACAAAC	GGCAAACTGT
40	1901	TTTTCAGCGG	CAGACCGACA	CCGCACGCCT	ACAATCATTT	AGGAAGCGGG
	1951	TGGTCAAAAA	TGGAAGGTAT	CCCACAAGGA	GAAATCGTGT	GGGACAACGA
	2001	CTGGATCNAC	CGCACGTTTA	AAGCGGAAAA	TTTCCATATT	CAGGGCGGGC
	2051	AGGCGGTGAT	TTCCCGCAAT	GTTGCCAAAG	TGGAAGGCGA	TTGNCATTTG
	2101	AGCAATCAGC	CCCAAGCAGT	TTTTGGTGTG	GCACCGCATC	AAAGCCATAC
	2151	AATCTGTACA	CGTTCGGACT	GGACNGGTCT	GACAAATTGT	GTCGAANAAA
45	2201	NCATTACCGA	GCATAAAGTG	ATTGCTTCAT	TGACTAAGAC	NGACNTNAGC
	2251	GGCANTGTNA	GNCTNNCCNA	TNACGNTNNT	TNAAANCTCN	CNNGGNTGTC
	2301	NNCACTNAAN	GGCAATCTTA	GTGCAATGCG	CGATACACGT	TATACAGTCA
	2351	GCCACAACGC	CACCCAAAC	GGCAACCTTA	GCCTCGTGGG	CAATGCCCAA
	2401	GCAACATTTA	ATCAAGCCAC	ATTAAACGGC	AACNCATCGG	NTTCGGGCAA
50	2451	TGCTTCATTT	AATCTAAGCA	ACAACGCCGC	ACAAAACGGC	AGTCTGACGC
	2501	TTTCCGACAA	CGCTAAGGCA	AACGTAAGCC	ATTCCGCACT	CAACGGCAAT
	2551	GTCTCCCTAG	CCGATAAGGC	AGTATTCCAT	TTTGAAAACA	GCCGCTTTAC
	2601	CGGACAACTC	AGCGGCAGCA	AGGANACAGC	ATTACACTTA	AAAGACAGCG
	2651	AATGGACGCT	GCCGTCAGGC	ACGGAATTAG	GCAATTTAAA	CCTTGACAAC
55	2701	GCCACCATTG	CACTCAATTC	CGCCTATCGC	CACGATGCTG	CAGGCGCGCA
	2751	AACCGGCAGN	GTGTCAGACA	CGCCGCGCGC	CCGTTCCGCG	CGTTCCCTAT
	2801	TATCCGTTAC	ACCGCCAAT	TCGGTAGAAT	CCCGTTTCAA	CACGCTGACG
	2851	GTAACCGGCA	AATTGAACNG	TCAAGGAACA	TTCCGCTTTA	TGTCGGAAC
	2901	CTTCGGCTAC	CGAAGCGACA	AATTGAAGCT	GGCGGAAAGT	TCCGAAGGNA
60	2951	CTTACACCTT	GGCGGTCAAC	AATACCGGCA	ACGAACCCGT	AAGCCTCGAT
	3001	CAATTGACGG	TAGTGGAAGG	GAAAGACAAC	AAACCGCTGT	CCGAAAACCT
	3051	TAATTTACCC	CTGCAAAACG	AACACGTCGA	TGCCGCGCGC	TGGCGTTACC
	3101	AACTCATCCG	CAAGACGGC	GAGTTCCGCC	TGCATAATCC	GGTCAAAGAA
	3151	CAAGAGCTTT	CCGCAAAACT	CGGCAAGGCA	GAAGCCAAAA	AACAGGCGGA
65	3201	AAAAGACAAC	GCGCAAAGCC	TTGACGCGCT	GATTGCGGCC	GGGCGCGATG
	3251	CCGCCGAAAA	GACAGAAAGC	GTTGCCGAAC	CGGCCCGGCN	GGCAGGCGGG
	3301	GAAATGTGCG	GCATTATGCA	GGCGGAGGAA	GAGAAAAAAC	GGGTGCAGGC
	3351	GGATAAAGAC	AGCGCNTTGG	CGAAACAGCG	CGAAGCGGAA	ACCCGGCCGG
	3401	NTACCACCGC	CTTCCCCCGC	GCCCCGNCGC	CCCCCGGGGA	TTTGGCCGAA
	3451	CCGCAGCCCC	AACCGCAACC	TCAACCCCAA	CCGCAGCGCG	ACCTGATNAG
70	3501	CCGTTATGCC	AATAGCGGTT	TGAGTGAATT	TTCCGCCACG	CTCAACAGCG
	3551	TTTTCCGCGT	ACAGGACGAA	TTGGACCGCG	TGTTTGCCGA	AGACCGCCGC

-365-

3601 AACGCNGTTT GGACAAGCNG CATCCGGNAC ACCAAACACT ACCGTTTCGCA
 3651 AGATTTCCGC GCCTACCGCC AACAAACCGA CCTGCGCCAA ATCGGTATGC
 3701 AGAAAAACCT CGGCAGCGGG CGCGTCGGCA TCCTGTTTTC GCACAACCGG
 5 3751 ACCGAAAACA NCTTCGACGA CGGCATCGGC AACTCGGCAC GGCTTGCCCA
 3801 CGGCGCCGTT TTCGGGCAAT ACGGCATCGG CAGGTTTCGAC ATCGGCATCA
 3851 GCACGGGCGC GGGTTTTAGC AGCGGCANTC TINTCAGACGG CATCGGAGGC
 3901 AAAATCCGCC GCCCGGTGCT GCATTACGGC ATTCAGGCAC GATACCGCGC
 3951 CGGTTTCGGC GGATTTCGGCA TCGAACCGTA CATCGGCGCA ACGCGCTATT
 10 4001 TCGTCCAAAA AGCGGATTAC CGCTACGAAA ACGTCAATAT CGCCACCCCC
 4051 GGTCTTGCCT TCAACCGNTA CCGNGCGGGC ATTAAGGCAG ATTATTCTAT
 4101 CAAACCGGCG CAACACATNT CCATCACNCC TTATTTNAGC CTGTCCTATA
 4151 CCGATGCCCG TTCGGGCAAA GTCCGAACAC GCGTCAATAC CGCNGTATTG
 4201 GCTCAGGATT TCGGCAAAAC CCGCAGTGGC GAATGGGGCG TAAACGCCGA
 4251 AATCAAAGGT TTCACGCTGT CCNTCCACGC TGCCGCGGCC AAAGNCCCG
 15 4301 AACTGGAAGC GCAACACAGC GCGGGCATCA AATTAGGCTA CCGCTGGTAA

This encodes a protein having amino acid sequence <SEQ ID 652>:

1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN
 51 YQYYRDFAEK KGKFAVGAKD IEVYNKKGEL VGKSMKAPM IDFSVVSRRNG
 20 101 VAALVGDQYI VSAHNGGYN NVDFGAEGXN PDQHRFSYQI VKRNNYKPPN
 151 SHPYNGDXHM PRLHKFVTD AEPVEMTSDMR GNTYSDKEY PERVRIGSGH
 201 HYWRYDDDKH GDLSYSGAWL IGGNTHMQGW GNGGVXSLSG DVRHANDYGP
 251 MPIAGAAGDS GSPMFIYDKT NNKWLNGVL QTGYPSGRE NGFQLIRKDW
 301 FYDDIYRGDT HTVXFEPNSN GHFSFTSNNN GTGTVTETNE KVSNPKLKVQ
 351 TVRLFDES LN ETDKEPVYAA GGVNQYRPR NNGENLSFID YNGKILSLN
 25 401 NINQAGGLY FEGDFTVSPE NNETWQAGV HISEDSTVTW KLVGVANDRL
 451 SKIGKGT LHV QAKGENQGS SVGDGT VILD QOADDKGGKQ AFSEIGLXSG
 501 RGTVQLNADN QFNPDKLYFG FRGRLDLNG HSLSFHRIQ TDEGAMIXXH
 551 NATTTSTVTI TGNESITQPS GKNINRLNYS KEIAYNGWFG EKDTTKTNGR
 601 LNLVYQPAE DRTXLLSGGT NLNGNITQTN GKLFSGRPT PHAYNHLGSG
 30 651 WSKMEGIPQ EIVWDNDWIX RTFKAENFHI QGGQAVISR N VAKVEGDHXL
 701 SNHAQAVFGV APHQSHITCT RSDWTGLTNC VEXXITDDKV IASLTKTDXS
 751 GXVXLXXXXX XXIXGXAXLX GNLSANGDTR YTVSHNATON GNLSLVGNAQ
 801 ATFNQATLNG NXSXSGNASF NLSNNAQNG SLTSLDNAK NVSHSALNGN
 851 VSLADKAVFH FENSRTGQL SGSKXTALHL KDSEWTLPSG TELGNLNLDN
 35 901 ATITLNSAYR HDAAGATGX VSDTPRRSR RSLSVTPPT SVESRFTLT
 951 VNGKLNQGT FRFMSELFY RSDKLKLAES SEGTYTLAVN NTGNEPVSLD
 1001 QLTVEGKDN KPLSENLFN LQNEHVDAGA WRYQLIRKDG EFR LHPVKE
 1051 QELSDKLGKA EAKKQAEKDN AQSLDALIAA GRDAEKTES VAEPARXAGG
 1101 ENVGIMQAE EKRRVQADKD SALAKQREAE TRPXTTAFPR ARXARRDLPO
 40 1151 POPQPOPQPO PORDLXSYA NSGLSEFSAT LNSVFAVQDE LDRVFAEDRR
 1201 NAVWTSXIRX TKHYRSQDFR AYRQOTDLRQ IGMQKNLGSV RVGILFSHNR
 1251 TENXFDDGIG NSARLAHGAV FGQYIGIRFD IGISTGAGFS SGXLSDDGIG
 1301 KIRRRVLHYG IQARYRAGFG GFGIEPYIGA TRYFVQADY RYENVNIATP
 1351 GLAFNRYRAG IKADYSFKPA QHXSITPYXS LSYTDAASGK VRTRVNTAVL
 45 1401 AQDFGKTRSA EWGVNAEIKG FTLSXHAAAA KGPQLEAQS AGIKLGYRW*

A transmembrane region is underlined.

ORF1-1 shows 86.3% identity over a 1462aa overlap with ORF1a:

		10	20	30	40	50	60
50	orfla.pep	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGIN	YQYYRDFAEK				
	orfl-1	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGIN	YQYYRDFAEK				
		10	20	30	40	50	60
55	orfla.pep	KGKFAVGAKDIEVYNKKGELVGKSMKAPMIDFSVVSRRNGVAALVGDQYIVSAHNGGYN					
	orfl-1	KGKFAVGAKDIEVYNKKGELVGKSMKAPMIDFSVVSRRNGVAALVGDQYIVSAHNGGYN					
		70	80	90	100	110	120
60	orfla.pep	NVDFGAEGXNPDQHRFSYQIVKRNNYKPDNS-HPYNGDXHMPRLHKFVTD AEPVEMTSDM					
	orfl-1	NVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTD AEPVEMTSDM					
65		130	140	150	160	170	180

BNSDOCID: <WO__992457BA2_1_>

-367-

		900	910	920	930	940
	orfla.pep	TELGNLNLNATITLNSAYRHDAAGAQ	TGXVSDTPRRRSRRS---	LLSVTPPTS	SVESRFN	
5	orfl-1	TELGNLNLNATITLNSAYRHDAAGAQ	TGXSATDAPRRRSRRSRLSVTPPTS	SVESRFN		
		900	910	920	930	940
	orfla.pep	950	960	970	980	990
10	orfl-1	950	960	970	980	990
		1010	1020	1030	1040	1050
15	orfla.pep	1010	1020	1030	1040	1050
	orfl-1	1010	1020	1030	1040	1050
		1070	1080	1090	1100	1110
20	orfla.pep	1070	1080	1090	1100	1110
	orfl-1	1070	1080	1090	1100	1110
		1130	1140	1150	1160	1170
25	orfla.pep	1130	1140	1150	1160	1170
	orfl-1	1130	1140	1150	1160	1170
30		1190	1200	1210	1220	1230
	orfla.pep	1190	1200	1210	1220	1230
	orfl-1	1190	1200	1210	1220	1230
35		1250	1260	1270	1280	1290
	orfla.pep	1250	1260	1270	1280	1290
	orfl-1	1250	1260	1270	1280	1290
40		1310	1320	1330	1340	1350
	orfla.pep	1310	1320	1330	1340	1350
	orfl-1	1310	1320	1330	1340	1350
45		1370	1380	1390	1400	1410
	orfla.pep	1370	1380	1390	1400	1410
	orfl-1	1370	1380	1390	1400	1410
50		1430	1440	1450		
	orfla.pep	1430	1440	1450		
	orfl-1	1430	1440	1450		

Homology with adhesion and penetration protein hap precursor of *H. influenzae* (accession number P45387)

Amino acids 23-423 of ORF1 show 59% aa identity with hap protein in 450aa overlap:

65	orfl	23	FXAAYLAICLSFGILPQAWAGHTYFG	INYYRDF	FAENKKGFAVGAKDIEVYNKKGELVG	82
			F +L C+S GI QAWAGHTYFGI+YQYYRDF	FAENKKGK F VGAK+IEVYNK+G+LVG		
	hap	6	FRLNFLTACVSLGIASQAWAGHTYFG	IDYQYYRDF	FAENKKGKFTVGAKNIEVYNKEGQLVG	65
	orfl	83	KSMTKAPMIDFSVSRNGVAALVGVQYIV	SAHNGGYNVDFGAEGXNIXDQXRTYKIV	142	
			SMTKAPMIDFSVSRNGVAALVG QYIVSAHNGGYN+VDFGAEG N DQ R TY+IV			
	hap	66	TSMTKAPMIDFSVSRNGVAALVGDQYIV	SAHNGGYNVDFGAEGRN-PDQHRFTYQIV	124	

-368-

5
10
15
20

orf1 143 KRNNYKAGTKGHPYGGDYHMPRLHKXVTDAPVEMTSYMDGRKYIDQNNYPDRVRIGAGR 202
KRNNY+A + HPY GDYHMPRLHK VT+AEPV MT+ MDG+ Y D+ NYP+RVRIG+GR
hap 125 KRNNYQAWERKHPYDGDYHMPRLHKFVTEAEPVGMTTNDGKVYADRENYPERVRIGSGR 184

orf1 203 QYWRSDDEPNRESSYHIA----- 222
QYWR+D+DE N SSY+++
hap 185 QYWRDDEETNVHSSYYVSGAYRYLTAGNTHQSGNGNGTVNLSGNVSPNHYGPLPTG 244

orf1 223 -----SGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGFQLVRKDWFYDEIFAGDTHSVF 277
SGSPMFIYDA+K++WLIN VLQTG+P+ G+ NGFQL+R++WFY+E+ A DT SVF
hap 245 GSKGDSGSPMFIYDAKKQWLINAVLQTGHFFGRGNGFQLIREEFYNEVLAVDTPSVF 304

orf1 278 --YEPRQNGKYSFNDNNGTGKIN-AKHEHNSLPNRLKTRTVQLFNVSLSETAREPVYHA 334
Y P NG YSF +N+GTGK+ + + + + TV+LFN SL++TA+E V A
hap 305 QRYIPPINGHYSFVSNNDGTGKLTITRPSKDGSKAKSEVGTVKLFNPSLNQTAKEHV-KA 363

orf1 335 AGGVNSYRPRLNNGENISFIDEGKGLILTSNINQAGGLYFQGDFTV-SPENNETWQGA 393
A G N Y+PR+ G+NI D+GKG L + +NINQAGGLYF+G+F V +NN TWQGA
hap 364 AAGYNIYQPRMEYGKNIYLGDOGKGTLTIENTNINQAGGLYFEGNFVVKGQNNITWQGA 423

orf1 394 GVHISEDSTVTWKVNGVANDRLSKIGKGT 423
GV I +D+TV WKV+ NDRLSKIG GTL
hap 424 GVSIGQDATVEWKVHNPENDRLSKIGIGTL 453

25 Amino acids 715-1011 of ORF1 show 50% aa identity with hap protein in 258aa overlap:

30
35
40

Orf1 41 DTRYTVSHNATQ-NGNXSLVXNAQATFNQ-ATLNGNTSASGNASFNLSHAVQNGSLTLS 98
DT+ S TQ NG+ +L NA + A LNGN + ++ F LS++A Q G++ LS
hap 733 DTKVINSIPITQINGSINLTNNATVNIHGLAKLNGNVTILIDHSOFTLSNNATQGNIKLS 792

orf1 99 GNAKANVSHSALNGNVSADKAVFHFESSRFTGQISGGKDTALHLKDSWTLPSPGKELGN 158
+A A V+++ LNGN L D A F ++S F QI G KDT + L+++ WT+PS L N
hap 793 NHANATVNNATLNGNVHLTDSAQFSLKNSHFWHQIQGDKDITVLEENATWTMPSDTTLQN 852

orf1 159 LNLDNATITLNSAYRHDAAGAQTGSATDAPXXXXXXXXXXLLXVTPPTSVESEFNTLTVN 218
L L+N+T+TLNSAY + S+ +AP L T PTS E RNTLTVN
hap 853 LTLNNSTVTLSAY-----SASSNNAPRHRS-----LETETTPTSAEHRENTLTVN 899

orf1 219 GKLNQGTFRFMSELFGRYSDKLKLAESSEGTYYTLAVNNTGNEPASLEQLTVVEGKDNKP 278
GKL+GQGTFF+ F S LFGY+SDKLKL+ +EG YTL+V NTG EP +LEQLT++E DNKP
hap 900 GKLSGQGTFFQFTSSLFYKSDKLKLSNDAEGDYTLVSRNTGKEPVTLEQLTLIESLDNKP 959

orf1 279 LSENLFNLQNEHVDAGA 296
LS+ L FTL+N+HVDAGA
hap 960 LSDKLKFTLENDHVDAGA 977

45 Amino acids 1192-1450 of ORF1 show 41% aa identity with hap protein in 259aa overlap:

50
55
60
65

Orf1 1 LDRVFAEDRRNAVWTSIGIRDTKHYRSQDFRAYRQOTDLRQIGMQKNLGSGRVGLFSHNR 60
LDR+F + ++AVWT+ +D + Y S FRAY+Q+T+LRQIG+OK L +GR+G +FSH+R
hap 1135 LDRLFVDQAQSAVWTNIAQDKRRYDSDAFRAYQKKTNLQIGVQKALANGRIGAVFSHSR 1194

orf1 61 TENTFDDGIGNSARLAHGAVFGQYGIDRFYXXXXXXXXXXXXXXXXXIGKXRRRVLHYG 120
++NTFD+ + N A L + F QY K R+ ++YG
hap 1195 SDNTFDEQVKNHATLTMMSGFAQYQWGDLOFGVNVGTGISASKMAEEQSRKIHRAINYG 1254

orf1 121 IQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYRAGIKADYSFKPA 180
+ A Y+ G GI+P+ G RYF+++ +Y+ E V + TP LAFNRY AGI+ DY+F P
hap 1255 VNASYQFRLGQLGIQPYFGVNRIFYERENYQSEEVVRVKTPLAFNRYNAGIRVDYFTPT 1314

orf1 181 QHISITPYLSLSYTDASGKVRTRVNTAVLAQDFGKTRSAEWGVNAEIKGFTLSLHAAAA 240
+IS+ PY ++Y D ++ V+T VN VL Q FG+ E G+ AEI F +S + +
hap 1315 DNISVKPYFFVNVYDVSNANVQTTVNLTVLQPPFGRYWQKEVGLKAEILHFQISAFISKS 1374

orf1 241 KGPQLEAQSAGIKLGYRW 259
+G QL Q + G+KLGYRW
hap 1375 QGSQLGKQNVGVKLGWR 1393

Homology with a predicted ORF from *N.gonorrhoeae*

The blocks of ORF1 show 83.5%, 88.3%, and 97.7% identities in 467, 298, and 259 aa overlap, respectively with a predicted ORF (ORF1ng) from *N.gonorrhoeae*:

5	orf1.pep	MKTTDKRTTETHRKAPKTGRIRFXAAYLAICLSFGILPOAWAGHTYFGINYQYYRDFAE	60
	orf1ng	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDFAE	60
10	orf1.pep	KGKFAVGAKDIEVYNKKGELVGKSMKAPMIDFSVSRNGVAALVGQYIVSVAHNGGYN	120
	orf1ng	KGKFAVGAKDIEVYNKKGELVGKSMKAPMIDFSVSRNGVAALAGDQYIVSVAHNGGYN	120
15	orf1.pep	NVDFGAEGXNIXDQXRXYKIVKRNNYKAGTKGHPYGGDYHMPRLHKKVTDAPVEMTSY	180
	orf1ng	NVDFGAEGSN-PDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKKFVTDAPVEMTSY	179
20	orf1.pep	MDGRKYIDQNNYPDRVRIGAGRQYWSDEDEPNRESSYHIAS-----	223
	orf1ng	MDGWKYADLNKYPDRVRIGAGRQYWSDEDEPNRESSYHIASAYSWLVGGNTFAQNGSG	239
25	orf1.pep	-----GSPMFIYDAQKQKWLINGVLOTGNPYIGKSNG	255
	orf1ng	GGTVNLGSEKIKHSPYGFLLPTGSGSFGDSGSPMFIYDAQKQKWLINGVLOTGNPYIGKSNG	289
30	orf1.pep	FOLVRKDWFYDEIFAGDTHSVFYEPHQNGKYSEFDDNNGTGKINAKHEHNSLPNRLKTRT	315
	orf1ng	FOLVRKDWFYDEIFAGDTHSVFYEPHQNGKYFFDNNNGAGKIDAKHKKHSLPYRLKTRT	359
35	orf1.pep	VQLFNVSLSETAREPVYHAAGGVNSYRPLNNGENISFIDEGKGELILTSNINQAGAGGLY	375
	orf1ng	VQLFNVSLSETAREPVYHAAGGVNSYRPLNNGENISFIDKGGKELILTSNINQAGAGGLY	
40	orf1.pep	FOGDFTVSPENNETWQAGAGVHISEDSTVTWKVNGVANDRLSKIGKGT	422
	orf1ng	FEGNFTVSPKNNETWQAGAGVHISDSTVTWKVNGVANDRLSKIGKGTLLVQAKGENQGSV	479
45	orf1.pep	// DKVTASLTKTDISGNVDLADHAHLNLTGLA	744
	orf1ng	FGVAPHQSHTICTRSDWTGLTSCTEKITDDKVIASLSKTDVRGNVSLADHAHLNLTGLA	774
50	orf1.pep	TLNGNLSANGDTR-YTVSHNATQNGNXSLVXNAQATFNQATLNGNTSASGNASFNLSNHA	803
	orf1ng	TFNGNL-VQAETRTIRLRANATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSNNA	833
55	orf1.pep	VQNGSLTSLGNAKANVSHSALNGNVSLADKAVHFESSRFTGQISGGKDTALHLKDSEWT	863
	orf1ng	VQNGSLTSLDNANAVSHSALNGNVSLADKAVHFENSRTGKISGGKDTALHLKDSEWT	893
60	orf1.pep	LPSGXLGNLNDNATITLNSAYRHDAAQAQTSATDAPRRRSRRSRLXVTPPTSVE	923
	orf1ng	LPSGTELGNLNDNATITLNSAYRHDAAQAQTSAAAPRRRSRRS---LLSVTPPTSVE	950
65	orf1.pep	SRFNTLTVNGKLNQGTFRFMSELFYGRSDKLKLAESSEGTYYTLAVNNTGNEPASLEQLT	983
	orf1ng	SRFNTLTVNGKLNQGTFRFMSELFYGRSGKLKLAESSEGTYYTLAVNNTGNEPVSLLEQLT	1010
70	orf1.pep	VVEGKDNKPLSENLFNLQNEHVDAGAW	1011
	orf1ng	VVEGKDNTPLENLNLQNEHVDAGAWRYQLIRKDGELFRLHNPVKEQELSDKLKGAGET	1070
75	orf1.pep	// LDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1211
	orf1ng	PQRDLISRYANSGLSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTKHYRSQDFR	1239
80	orf1.pep	AYRQOTDLRQIGMQKNLGSGRVGLFSHNRTENTFDDGIGNSARLAHGAVFQYQIDRFY	1271
	orf1ng	AYRQOTDLRQIGMQKNLGSGRVGLFSHNRTGNTFDDGIGNSARLAHGAVFQYQIGRFD	1299


```

orfl.pep      IGISAGAGFSSGSLSDGIGKXRRRLVHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY 1331
               |||||
orflng        IGISAGAGFSSGSLSDGIRGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADY 1359

5  orfl.pep      RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDASGKVRTRVNTAVL 1391
               |||||
orflng        RYENVNIATPGLAFNRYRAGIKADYSFKPAQHISITPYLSLSYTDASGKVRTRVNTAVL 1419

10 orfl.pep      AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKPQLEAQSAGIKLGYRW 1440
               |||||
orflng        AQDFGKTRSAEWGVNAEIKGFTLSLHAAAAGKPQLEAQSAGIKLGYRW 1468

```

The complete length ORF1ng nucleotide sequence was identified <SEQ ID 653>:

```

1  ATGAAAACAA CCGACAAACG GACAACCGAA ACACACCGCA AAGCCCTTAA
15 51 AACCGGCCGC ATCCGCTTCT CGCCCGCTTA CTTAGCCATA TGCCTGTCGT
101 TCGGCATTCT GCCCAAGCC CGGGCGGGAC ACATTATTT CGGCATCAAC
151 TACCAATACT ATCGCGACTT TGCCGAAAT AAAGGCAAGT TTGCAGTCGG
201 GCGGAAAGAT ATTGAGGTTT ACAACAAAA AGGGGAGTTG GTCGGCAAA
251 CGATGACGAA AGCCCGATG ATTGATTTT CTGTGGTATC GCGTAACGGC
301 GTGGCGGCAT TGGCGGGCGA TCAATATATT GTGAGCGTGG CACATAACGG
20 351 CGGTATAAC AATGTTGATT TTGGTGCGGA GGAAGCAAT CCGATCAGC
401 ACGGCTTTTC TTACCAATT GTGAAAAGAA ATAATTATA AGCAGGACT
451 AACGGCCATC CTTATGGCGG CGATTATCAT ATGCCGCGTT TGCACAAATT
501 TGTACAGAT GCAGAACCTG TTGAGATGAC CAGTTATATG GATGGGTGGA
551 AATACGCTGA TTTAAATAAA TACCCTGATC GTGTCGAAT CGGAGCAGGC
25 601 AGACAATATT GCGGCTCTGA TGAAGACGAA CCCAATAACC GCGAAAGTTC
651 ATATCATATT GCAAGCGCAT ATTCTTGGCT CGTCGGTGGC AATACCTTTG
701 CACAAAATGG ATCAGGTGGT GGCACAGTCA ACTTAGGTAG CGAAAAAATT
751 AAACATAGCC CATATGGTTT TTTACCAACA GGAGGCTCAT TTGGCGACAG
801 TGGCTCACCA ATGTTTATCT ATGATGCCCA AAAGCAAAAG TGGTTAATTA
30 851 ATGGGGTATT GCAACAGGC AACCCCTATA TAGGAAAAG CAATGGCTTC
901 CAGCTAGTTC GTAAAGATTG GTTCTATGAT GAAATCTTTG CTGGAGATAC
951 CCATTCACTA TTCTACGAAC CACATCAAAA TGGGAAATAC TTTTTTAAAC
1001 ACAATAATAA TGGCGCAGGA AAAATCGATG CCAACATAA AACTATTCT
1051 CTACCTTATA GATTAATAAC ACGAACCGTT CAATTGTTTA ATGTTCTTT
35 1101 ATCCGAGACA GCAAGAGAAC CTGTTTATCA TGCTGCAGGT GGGGTCAACA
1151 GTTATCGACC CAGACTGAAT AATGGAGAAA ATATTTCCTT TATTGACAAA
1201 GGAAAAGGTG AATTGATACT TACCAGCAAC ATCAACCAAG GCGCGGGCGG
1251 TTTGTATTTT GAGGGTAATT TTACGGTCTC GCCTAAAAAC AACGAAACGT
1301 GGCAAGGCGC GGGCGTTCAT ATCAGTGATG GCAGTACCGT TACTTGAAA
40 1351 GTAAACGGCG TGGCAAAACGA CCGCCTGTCC AAAATCGGCA AAGGCACGCT
1401 GCTGGTTCAA GCCAAAGGGG AAAACCAAGG CTCGGTCAGC GTGGCGCAGC
1451 GTAAAGTCAT CTTAGATCAG CAGGCGGACG ATCAAGGCAA AAAACAAGCC
1501 TTTAGTGAAT TCGGCTTGGT CAGCGGCAGG GGGACGGTGC AACTGAATGC
1551 CGATAATCAG TTCAACCCCG ACAAACTCTA TTTCCGCTTT CGCGCGGAC
45 1601 GTTTGGATTT GAACGGGCAT TCGCTTTCGT TCCACCGCAT TCAAAATACC
1651 GATGAAGGGG CGATGATGTT CAACCACAAT CAAGACAAAG AATCCACCGT
1701 TACCATTACA GGCAATAAAG ATATTACTAC AACCGCAAT AACAACTCT
1751 TGGATAGCAA AAAAGAAATT GCCTACAACG GTTGGTTTGG CGAGAAAGAT
50 1801 GCAACCAAAA CGAACGGGCG GCTCAATCTG AATTACCAAC CGGAAGAAGC
1851 GGATCGCACT TACTGCTTT CCGGCGGAAC AAATTTAAAC GGCAATATCA
1901 CGCAACAAA CGGCAAACTG TTTTTCAGCG GCAGACCGAC ACCGACGCGC
1951 TACAATCATT TAGGAAGCGG GTGGTCAAAA ATGGAAGGTA TCCACAAGG
2001 AGAAATCGTG TGGGACAACG ATTGATCGA CCGCACATTT AAAGCGGAAA
2051 ACTTCCATAT TCAGGGCGGA CAAGCGGTGG TTTCCGCAA TGTGCCAAA
55 2101 GTGGAAGGCG ATTGGCATT TAAAGCAATCAG GCCAAGCAG TTTTCGGTGT
2151 CGCACCGCAT CAAAGCCACA CAATCTGTAC ACGTTCGGAC TGGACGGGTC
2201 TGACAAGTTG TACCGAAAAA ACCATTACCG ACGATAAAGT GATTGCTTCA
2251 TTGAGCAAGA CCGACATCAG AGGCAATGTC AGCCTTGCG ATCAGCTCA
2301 TTTAAATCTC ACAGGACTTG CCACACTCAA CGGCAATCTT AGTGCAGGCG
60 2351 GAGACACGCA CTATACGGTT ACGCGCAACG CCACCAAAA CGGCAACCTC
2401 AGCCTCGTGG GCAATGCCCA AGCAACATTT AATCAAGCCA CATTAAACGG
2451 CAACACATCG GCTTCGGACA ATGCTTCATT TAATCTAAGC AACCAACCGC
2501 TACAAAACGG CAGTCTGACG CTTTCCGACA ACGCTAAGGC AAACGTAAGC
2551 CATTCCGCAC TCAACGGCAA TGTCTCCCTA GCCGATAAGG CAGTATTCCA
65 2601 TTTTGAAAAC AGCCGCTTTA CCGGAAAAAT CAGCGGCGGC AAGGATACGG
2651 CATTACACTT AAAAGACAGC GAATGGACGC TGCCGTCGGG CACGGAATTA
2701 GGCAATTTAA ACCTTGACAA CGCCACCATT AACTCAATT CCGCTATCG
2751 ACACGATGCG GCAGGCGCGC AAACCGGCAG TGCGGCAGAT GCGCGCGGCC
2801 GCCGTTCCGC CCGTTCCCTA TTATCCGTTA CGCCGCCAAC TTCGGCAGAA
70 2851 TCCCGTTTCA ACACGCTGAC GGTAAACGGC AAATTGAACG GTCAGGGAAC

```

2901 ATTCCGCTTT ATGTCGGAAC TCTTCGGCTA CCGCAGCGGC AAATTGAAGC
 2951 TGGCGGAAAG TTCCGAAGGC ACTTACACCT TGGCTGTCAA CAATACCGGC
 3001 AACGAACCCG TAAGTCTCGA GCAATTGACG GTAGTGGAAG GAAAAGACAA
 3051 CACACCGCTG TCCGAAAATC TTAATTTTAC CCTGCaaaAc gaacacgtcg
 3101 atgccggcgc atggCGTTAT CAGCTTATCC gcaaagacgG CGAGTTCCGc
 3151 CTGCATAATC CCGTCAAAGA ACAAGAGCTT TCCGACAAAC TCGGCAAGc
 3201 gggagaaACA GAggcccgcT TGACGGCAAA ACAGGCacaA CTTGCCGCCA
 3251 AAcacaggc ggaAAAGAC AACgcgcaaa gccttgAcgc gctgattgcg
 3301 gCggggcgca atgccaccga AAAGGCAGaa agtgtgcccg aaccgGCCCG
 3351 GCAGGCAGGC GGGGAAAAtg ccgGCATTAT GCAGGCGGAG GAAGAGAAAA
 3401 AACGGGTGCA GCGCGATAAA GACACCGCCT TGGCGAAACA GCGCGAAGCG
 3451 GAAACCCGGC CGGCTACCAC CGCCTTCCCC CGCGCCCGCC GCGCCGCGCG
 3501 GGATTGCGCG CAACCGCAGC CCAACCGCA ACCCAACCGC CAGCGCGACC
 3551 TGATCAGCCG TTATGCCAAT AGCGGTTTGA GTGAATTTTC CGCCACGCTC
 3601 AACAGCGTTT TCGCGGTACA GGACGAATTG GACCGCGTGT TTGCCGAAGA
 3651 CCGCCGCAAC GCCGTTTGA CAAGCGGCAT CCGGGACACC AAACACTACC
 3701 GTTCGCAAGA TTTCCGCGCC TACCGCCAAC AAACCGACCT GCGCCAAATC
 3751 GGTATGCAGA AAAACCTCGG CAGCGGGCGC GTCGGCATCC TGTTTTCGCA
 3801 CAACCGGACC GGAAACACCT TCGACGACGG CATCGGCAAC TCGGCACGGC
 3851 TTGCCACGG TGCCGTTTTT GGGCAATACG GCATCGGCAG GTTCGACATC
 3901 GGCATCAGCG CGGGCGCGGG TTTTAGTAGC GGCAGCCTTT CAGACGGCAT
 3951 CAGAGGCAAA ATCCGCGCGC GCGTGCTGCA TTACGGCATT CAGGCAAGAT
 4001 ACCGCGCAGG TTTCCGCGGA TTCGGCATCG AACCGCACAT CGGCGCAACG
 4051 CGCTATTTCG TCCAAAAGC GGATTACCGA TACGAAAACG TCAATATCGC
 4101 CACCCCGGGC CTGTCATTCA ACCGCTACCG CGCGGGCATT AAGGCAGATT
 4151 ATTCAATCAA ACCGGCGCAA CACATTTCGA TCACGCCTTA TTTGAGCCTG
 4201 TCCTATACCG ATGCCGCTTC CGGCAAAGTC CGAACGCGCG TCAATACCGC
 4251 CGTATTGGCG CAGGATTTCG GCAAACCCG CAGTGCGGAA TGGGGCGTAA
 4301 ACGCCGAAAT CAAAGGTTTC ACGCTGTCCC TCCACGCTGC CGCCGCCAAG
 4351 GGGCCGCAAT TGGAAGCGCA GCACAGCGCG GGCATCAAT TAGGCTACCG
 4401 CTGGTAA

This is predicted to encode a protein having amino acid sequence <SEQ ID 654>:

1 MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPOA RAGHTYFGIN
 35 51 YQYYRDFAEK KGFVAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRRNG
 101 VAALAGDQYI VSAHNGGYN NVDFGAEGSN PDQHRFSYQI VKRNNYKAGT
 151 NGHPYGGDYH MPRLLHKFVD AEPVEMTSYM DGWKYADLNK YPDRVRIGAG
 201 RQYWRSEDEE PNNRESSYHI ASAYSWLVG GNTFAQNGSGG GTVNLGSEKI
 251 KHSPYGFLLPT GGSFGDGSGP MEIYDAQKQK WLINGVLOTG NPYIGKSNGF
 301 QLVRKDWFYD EIFAGDTHSV FYEPHQNGKY FENDNNNGAG KIDAKHKHYS
 351 LPYRLKTRTV QLFNVLSLET AREPVYHAAG GVNSYRPRLN NGENISFIDK
 401 GKGLILTSN INQGAGGLYF EGNFTVSPKN NETWQAGGVH ISDGSTVTWK
 451 VNGVANDRLS KIGKGTLLVQ AKGENQGSVS VGDGKVILDQ QADDQGGKQA
 501 FSEIGLVSGR GTVQLNADNQ FNPDKLYFGF RGGRLDLNGH SLSFHRIONT
 551 DEGAMIVNHN QDKESTVTIT GNKDITTTGN NNNLDSKKEI AYNWGFGEKD
 601 ATKTNGLNL NYPPEEADRT LLSGGTNLN GNITQTNGKL FFSGRPTPHA
 651 YNHLGSGWSK MEGIPQGEIV WDNDWIDRTF KAENFHIQGG QAVVSRNVAK
 701 VEGDWHLNSH AQAVFGVAPH QSHTICTRSD WTGLTSCTEK TITDDKVIAS
 751 LSKTDVRGNV SLADHAHLNL TGLATFNGNL VQAEIRTRIL RANATQNGNL
 801 SLVGNAQATF NQATLNGNTS ASDNASFNLS NNAVQNGSLT LSDNAKANVS
 851 HSALNGNVSL ADKAVFHFEN SRTGKISGG KDTALHLKDS EWTLPSTEL
 901 GNLNLDNATI TLNSAYRHDA AGAQTGSAAD APRRRSRRLS LSVTPPTSSE
 951 SRFNTLTVNG KLNGQGTFRF MSELFYGRSG KLKLAESSEG TYTLAVNNTG
 1001 NEPVSLQLT VVEGKDNTPL SENLNFLLQN EHVDAGAWRY QLIRKDGFEF
 1051 LHNVPVKEQEL SDKLGKAGET EAALTAKQAO LAAKQQAEDK NAQSLDALIA
 1101 AGRNATEKAE SVAEPARQAG GENAGIMQAE EEKRVQADK DTALAKQREA
 1151 ETRPATTAFF RARRARRDLP QPQPQPQPQ QRODISRYAN SGLSEFSATL
 1201 NSVFAVQDEL DRVFAEDRRN AVWTSGIRDT KHYRSQDFRA YRQQTDLRQI
 1251 GMQKNLGSGR VGILFSHNRT GNTFDDGIGN SARLAHGAVF GQYIGRFDI
 1301 GISAGAGFSS GSLSDGIRGK IRRRVLYHYI QARYRAGFGG FGIEPHIGAT
 1351 RYFVQKADYR YENVNIATPG LAFNRYRAGI KADYSFKPAQ HISITPYLSL
 1401 SYTDAASGKV RTRVNTAVLA QDFGKTRSAE WGVNAEIKGF TSLHAAAAAK
 1451 GPQLEAQHSA GIKLYRW*

Underlined and double-underlined sequences represent the active site of a serine protease (trypsin family) and an ATP/GTP-binding site motif A (P-loop).

65 ORF1-1 and ORF1ng show 93.7% identity in 1471 aa overlap:

-372-

5	orfl-1.pep	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQAWAGHTYFGINYQYYRDAEN
	orflng-1	MKTTDKRTTETHRKAPKTGRIRFSPAYLAICLSFGILPQARAGHTYFGINYQYYRDAEN
10	orfl-1.pep	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALVGDQYIVSVAHNGGYN
	orflng-1	KGKFAVGAKDIEVYNKKGELVGKSMTKAPMIDFSVVS RNGVAALAGDQYIVSVAHNGGYN
15	orfl-1.pep	NVDFGAEGRNPDQHRFTYKIVKRNNYKAGTKGHPYGGDYHMPRLHKFVTD AEPVEMTSYM
	orflng-1	NVDFGAEGSNPDQHRFSYQIVKRNNYKAGTNGHPYGGDYHMPRLHKFVTD AEPVEMTSYM
20	orfl-1.pep	DGRKYIDQNNYPDRVRIGAGRQYWRSD EEPNNRESSYHIASAYS WLVGNTFAQNGSGG
	orflng-1	DGWKYADLNKY PDRVRIGAGRQYWRSD EEPNNRESSYHIASAYS WLVGNTFAQNGSGG
25	orfl-1.pep	GTVNLGSEKIKHSPYGF LPTGGSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGF
	orflng-1	GTVNLGSEKIKHSPYGF LPTGGSGSPMFIYDAQKQKWLINGVLQTGNPYIGKSNGF
30	orfl-1.pep	QLVRKDWFYDEIFAGDTHSVFYEP RONGKYSFNDDNNGTGKINAKHEHNSLPNRLKTRTV
	orflng-1	QLVRKDWFYDEIFAGDTHSVFYEP HQNGKYFFNDNNNGAGKIDAKHKHYS LPYRLKTRTV
35	orfl-1.pep	QLFNVSLSETAREPVYHAAGGVNSYR PRLNNGENISFIDEGKGELILTSNINQAGAGGLYF
	orflng-1	QLFNVSLSETAREPVYHAAGGVNSYR PRLNNGENISFIDKGKGELILTSNINQAGAGGLYF
40	orfl-1.pep	QGDFTVSPENNETWQAGVHISEDSTV TWKVN GVANDRLSKIGKGLTHVQAKGENQGSIS
	orflng-1	EGNFTVSPKNNETWQAGVHISDGSTV TWKVN GVANDRLSKIGKGLTLLVQAKGENQGSVS
45	orfl-1.pep	VGDGTVILDQQADDKGGKQAFSEI GLVSGRGT VQLNADNQFNPDKLYFGFRGRLDLNGH
	orflng-1	VGDGKVIDLQQADDQGGKQAFSEI GLVSGRGT VQLNADNQFNPDKLYFGFRGRLDLNGH
50	orfl-1.pep	SLSFHRIQNTDEGAMIVNHNQDKESTV TITGNKD IATTGNNNSLDSKKEIAYNGWFGEKD
	orflng-1	SLSFHRIQNTDEGAMIVNHNQDKESTV TITGNKD ITTGNNNNLDSKKEIAYNGWFGEKD
55	orfl-1.pep	TTKTNGRLNLVYQPAEDRTL LLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLNDHWSQ
	orflng-1	ATKTNGRLNLNYQPEEADRTL LLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLGSGWSK
60	orfl-1.pep	KEGI PRGEI VWDNDWINRTFKAENFQIKGGQAVVSRNVAKVKGDWHL SNHAQAVFGVAPH
	orflng-1	MEGI PQGEI VWDNDWIDRTFKAENFHIQGGQAVVSRNVAKVEGDWHL SNHAQAVFGVAPH
65	orfl-1.pep	MEGI PQGEI VWDNDWIDRTFKAENFHIQGGQAVVSRNVAKVKGDWHL SNHAQAVFGVAPH
	orflng-1	MEGI PQGEI VWDNDWIDRTFKAENFHIQGGQAVVSRNVAKVKGDWHL SNHAQAVFGVAPH
70	orfl-1.pep	MEGI PQGEI VWDNDWIDRTFKAENFHIQGGQAVVSRNVAKVKGDWHL SNHAQAVFGVAPH
	orflng-1	MEGI PQGEI VWDNDWIDRTFKAENFHIQGGQAVVSRNVAKVKGDWHL SNHAQAVFGVAPH

```

              730      740      750      760      770      780
orf1-1.pep  QSHTICTRSDWTGLTNCVEKTTIDDKVIASLTKTDISGNVDLADHAHLNLTGLATLNGNL
              |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf1ng-1    QSHTICTRSDWTGLTSCFEKTTIDDKVIASLSKTDIRGNVLADHAHLNLTGLATLNGNL
              |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

10

```

              790      800      810      820      830      840
orf1-1.pep    SANGDTRYTVSHNATQNGNLSLVGNAQATFNQATLNGNTSASGNASFNLSDHAVONGSLT
               ||::|||::|||::|||||::|||||::|||||::|||||::|||::|
orf1ng-1      SAGGDTHYTVTRNATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLSSNAVONGSLT
              790      800      810      820      830      840

```

15

```

      850      860      870      880      890      900
orf1-1.pep  LSGNAKANVSHSALNGNVSLADKAVFHFESSRFTGQISGGKDTALHLKDSEWTLPSGTTEL
              |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf1ng-1     LSDNAKANVSHSALNGNVSLADKAVFHFENSRTGKISGGKDTALHLKDSEWTLPSGTTEL
              850      860      870      880      890      900

```

20

```

          910          920          930          940          950          960
orf1-1.pep  GNLNLNDNATITLNSAYRHAAGAQTGSATDAPRRRRRRRRSLLSVTPPTSVESRFNTLT
             |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf1ng-1    GNLNLNDNATITLNSAYRHAAGAQTGSAADAPRRRSR---RSLLSVTPPTSAESRFNTLT
          910          920          930          940          950

```

25

```

          970      980      990      1000      1010      1020
orf1-1.pep  VNGKLNQGQTFRFMSELFGRSDKCLKLAESSEGYTTLAVNNTGNEPASLEQLTVVEGKDN
             |||||
orf1ng-1    VNGKLNQGQTFRFMSELFGRSGKCLKLAESSEGYTTLAVNNTGNEPVSLEQLTVVEGKDN
          960      970      980      990      1000      1010

```

30

```

              1030      1040      1050      1060      1070
orf1-l.pep  KPLSENLNFTLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKA-----
              |||||
orf1ng-1    TPLSENLNFTLQNEHVDAGAWRYQLIRKDGEFRLHNPVKEQELSDKLGKAGETEAALTAK
              1020      1030      1040      1050      1060      1070

```

35

```

                                1080      1090      1100      1110      1120
orf1-1.pep  ----EAKKQAEKDNAQSLDALIAAGRDAVEKTESVAEPAEQAGGENVVGIMQAEEEKKRVO
              ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
orf1ng-1     QAQLAAKQAEKDNAQSLDALIAAGRNATEKAESVAEPAEQAGGENAGIMQAEEEKKRVO
              1080      1090      1100      1110      1120      1130

```

45

```

      1130      1140      1150      1160      1170      1180
orf1-1.pep ADKDTALAKQREAETRPATTAFPRARRARRDLPQLQPQPQPQPQRDLISRYANSGLSEFS
            |||||
orf1ng-1   ADKDTALAKQREAETRPATTAFPRARRARRDLPQPQPQPQPQRDLISRYANSGLSEFS
      1140      1150      1160      1170      1180      1190

```

50

```

1190      1200      1210      1220      1230      1240
orf1-1.pep  ATLNSVFAVQDELDRVFAEDRRNAVWTSGIRDTHKHYRSQDFRAYRQOTDLRQIGMQKNLG
            |||||
orf1ng-1    ATLNSVFAVQDELDRVFAEDRRNAVWTSGIRDTHKHYRSQDFRAYRQOTDLRQIGMQKNLG
            1200      1210      1220      1230      1240      1250

```

55

```

      1250      1260      1270      1280      1290      1300
orfl-1.pep  SGRVGILFSHNRTENTFDDGIGNSARLAHGAVFGQYGIDRFYIGISAGAGFSSGSLSDGI
              |||||
orflng-1    SGRVGILFSHNRTGNTFDDGIGNSARLAHGAVFGQYGIGRFDIGISAGAGFSSGSLSDGI
      1260      1270      1280      1290      1300      1310

```

60

```

1310      1320      1330      1340      1350      1360
orf1-1.pep  GGGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYR
              |||||
orf1ng-1    RGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGLAFNRYR
1320      1330      1340      1350      1360      1370

```

65

```

1370      1380      1390      1400      1410      1420
orf1-1.pep AGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
orf1ng-1    AGIKADYSFKPAQHISITPYLSLSYTDAAAGKVRTRVNTAVLAQDFGKTRSAEWGVNAEI
1380      1390      1400      1410      1420      1430

```

```

                                1430      1440      1450
orf1-1.pep  KGFTLSLHAAAAAGPQLEAQHSAGIKLGYRWX
              |||||
orf1ng-1    KGFTLSLHAAAAAGPQLEAQHSAGIKLGYRWX
              |||||
            1440      1450      1460

```

SCORES Init1: 1104 Initn: 4632 Opt: 2680
Smith-Waterman score: 5165; 55.7% identity in 1455 aa overlap

BNSDOCID: <WO__9924578A2_1_>

		600	610	620	630	640	650
	orf1ng-1.pep	EKDATKTNGRLNLNYQPEEADRTLLLSGGTNLNGNITQTNGKLFFSGRPTPHAYNHLGSG					
5	p45387	ETDKNKHNGRLNLNLIYKPTTEDRTLLLSGGTNLKGDTQTKGKLFFSGRPTPHAYNHLNKR	570	580	590	600	610
						620	
		660	670	680	690	700	710
10	orf1ng-1.pep	WSKMEGIPQGEIVWDNDWIDRTFKAENFHIQGGQAVVSRNVAKVEGDWHLNSHAQAVFGV					
	p45387	WSEMEGIPQGEIVWDHDWINRTFKAENFQIKGGSVAVSRNVSSIEGNWTVSNANATFGV	630	640	650	660	670
						680	
		720	730	740	750	760	770
15	orf1ng-1.pep	APHQSHTICTRSDWTGLTSCTEKTITDDKVIASLSKTDIRGNVSLADHAHLNLTGLATLN					
	p45387	VPNQNTICTRSDWTGLTTCQKVDLTDTKVINISIPKTQINGSINLTDNATANVKGGLAKLN	690	700	710	720	730
						740	
		780	790	800	810	820	830
20	orf1ng-1.pep	GNLSAGGDTHYTVTRNATQNGNLSLVGNAQATFNQATLNGNTSASDNASFNLNNAVQNG					
	p45387	GNVTL-----TNHSQFTLSNNATQIG	750			760	770
25		840	850	860	870	880	890
	orf1ng-1.pep	SLTLSDNAKANVSHSALNGNVSLADKAVFHFENSRTGKISGGKDTALHLKDSEWTLPSG					
	p45387	NIRLSDNSTATVDNANLNGNVHLTDSAQFSLKNSHFSHQIQGDKGTTVTLENATWTMPSD	780	790	800	810	820
30						830	
		900	910	920	930	940	950
	orf1ng-1.pep	TELGNLNLDNATITLNSAYRHDAAQAQTGSAADAPRRRSRSLSVTPPTSAESRFTLT					
35	p45387	TTLQNLTLNNSTITLNSAY-----SASSNTPRRRS---LETETPTSAEHRFTLT	840	850	860	870	
		960	970	980	990	1000	1010
40	orf1ng-1.pep	VNGKLNQGQTFRMSSELFGRSGKLKLAESSEGTYTLAVNNTGNPEVSLEQLTVVEGKDN					
	p45387	VNGKLSGGQTFQFTSSLFYKSDKLKLSNDAEGDYILSVRNTGKEPETLEQLTLVESKDN	880	890	900	910	920
						930	
		1020	1030	1040	1050	1060	1070
45	orf1ng-1.pep	TPLSENLFNLQNEHVDAGAWRYQLIRKDGFEFRLHNPVKEQLSDKLGKAGETEALTA					
	p45387	QPLSDKLFKFTLENDHVDAGALRYKLVKNDGEFRLHNPKEQLHNDLVRAEQAERTLEAK	940	950	960	970	980
						990	
		1080	1090	1100	1110	1120	1130
50	orf1ng-1.pep	QAQLAAKQQAQKDNQSLDALIAAGRNAT-EKAESVAEPARQAGGENAGIMQAEKEKRV					
	p45387	QVEPTAKTQTGEPKVRSRRAARAAPDFTLPDQSLNLALEAKQAE-LTAETQKSKAKTKKV	1000	1010	1020	1030	1040
						1050	
		1140	1150	1160	1170	1180	1190
55	orf1ng-1.pep	QADK---DTALAKQREAEATRPATTAFFRARRARRD-LPQPQPQPQPQDRLISRYANS					
	p45387	RSKRARVSDPLLDQSLFALEAVIDAPQQSEKDRLAQEEAEQ-RKQKDLISRYNSA	1060	1070	1080	1090	1100
60						1110	
		1200	1210	1220	1230	1240	1250
	orf1ng-1.pep	LSEFSATLNSVFAVQDELDRVFAEDRRNAVWTSIGIRDTHKHYRSQDFRAYRQQ-TDLRQIG					
65	p45387	LSELSATVNSMLSVQDELDRFLVDQAQSAVWNTNIAQDKRRYDSDAFRAYQQQKTNLRQIG	1120	1130	1140	1150	1160
						1170	
		1260	1270	1280	1290	1300	1310
	orf1ng-1.pep	MQKNLGSGRVGLFHSNRTGNTFDDGIGNSARLAHGAVFQYQIGRFDIGISAGAGFSSG					
70	p45387	VQKALANGRIGAVFHSRSNDTFDEQVKNHATLTMSGFAQYQWGDQLQFGVNVGTGISAS	1180	1190	1200	1210	1220
						1230	

-376-

		1320	1330	1340	1350	1360	1370
	orf1ng-1.pep	SLSDGIRGKIRRRVLHYGIQARYRAGFGGFGIEPHIGATRYFVQKADYRYENVNIATPGL					
5	p45387	1240	1250	1260	1270	1280	1290
		1380	1390	1400	1410	1420	1430
	orf1ng-1.pep	AFNRYRAGIKADYSFKPAQHISITPYLSLSYTDASGKVRTRVNTAVLAQDFGKTRSAEW					
10	p45387	1300	1310	1320	1330	1340	1350
		1440	1450	1460	1469		
15	orf1ng-1.pep	GVNAEIKGFTLSLHAAAAGPQLEAQHSAGIKLGYRWX					
	p45387	1360	1370	1380	1390		

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

20 Example 78

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 655>:

25	1	..AAGGTGTGGC	AATTGTGCGA	AGA.CCGCTG	CGTGCCGTCG	TGCCTGCCGA
	51	CAGTTTGTAA	CCGACCGCGC	AAAAATTGAA	CCTGTTTAAG	GCGGGTGCGG
30	101	CAACCATTTT	GTTTATGAA	GATCAAAATG	TCGTCAAAGG	TTTGCAGGAG
	151	CAGTCCCTG	CTTATGCCG	TAACCTCCCC	GTTGGGGCGg	ATCAGGCAAA
	201	CGCGATGGT	CAGTATGCCG	TTTGGACGAC	ACTTGCCGCG	GTCGGCGTAG
	251	GTGCAAACCT	GCAACATTAC	AATCCCTTGC	CCGATGCCGC	GATTGCCAAA
	301	GCGTGAATA	TCCCCGAAAA	CTGGTTGTTG	CGCGCACAAA	TGTTATCGG
	351	CGGTATTGAA	GGGGCGGCAG	GTGAAAAGAC	CTTGAACCC	GTTGCAGAAC
	401	GTTTGAAAGT	GTTGCGCGCA	TAA		

This corresponds to the amino acid sequence <SEQ ID 656; ORF6>:

35	1	..KVVQFVEXPL	RAVVPADSFE	PTAQKLNLFK	AGAATILFYE	DQNVVKGLQE
	51	QFPAYAANFP	VWADQANAMV	QYAVWTTLAA	VGVGANLQHY	NPLPDAAIAK
	101	AWNIPENWLL	RAQMVIGGIE	GAAGEKTFEP	VAERLKVFGA	*

Further sequence analysis revealed a further partial DNA sequence <SEQ ID 657>:

40	1	..CTGCGTGCCG	TCGTGCCTGC	CGACAGTTTT	GAACCGACCG	CGCAAAAATT
	51	GAACCTGTTT	AAGGCGGGTG	CGGCAACCAT	TTTGTTTTAT	GAAGATCAAA
	101	ATGTCGTCAA	AGGTTTGACG	GAGCAGTTCC	CTGCTTATGC	CGCTAATTC
	151	CCCGTTTGGG	CGGATCAGGC	AAACGCGATG	GTGCAGTATG	CCGTTTGGAC
	201	GACACTTGCC	GCGGTCGGCG	TAGGTGCAAA	CCTGCAACAT	TACAATCCCT
	251	TGCCCGATGC	GCGGATTGCC	AAAGCGTGGA	ATATCCCCGA	AAACTGGTTG
	301	TTGCGCGCAC	AAATGGTTAT	CGGCGGTATT	GAAGGGCGCG	CAGGTGAAAA
	351	GACCTTTGAA	CCCGTTGCAG	AACGTTTGAA	AGTGTTCGGC	GCATAA

This corresponds to the amino acid sequence <SEQ ID 658; ORF6-1>:

45	1	..LRVVPADSF	EPTAQKLNLF	KAGAATILFY	EDQNVVKGLQ	EQFPAYAANF
	51	PVWADQANAM	VQYAVWTTLA	AVGVGANLQH	YNPLPDAAIA	KAWNIPENWL
	101	LRAQMVIGGI	EGAAGEKTFE	PVAERLKVFG	A*	

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF6 shows 98.6% identity over a 140aa overlap with an ORF (ORF6a) from strain A of *N.meningitidis*:

10 20 30

-377-

```

      orf6.pep                                KVVQFVEXPLRAVVPADSFEPTAQKLNLFK
      orf6a      QIVEHAVLHTPSSFNSQSARVVVLFGEEDKVVQFVEDALRAVVPADSFEPTAQKLNLFK
                    40      50      60      70      80      90
5
      orf6.pep      AGAATILFYEDQNVVKGLEQFPAYAAFPVWADQANAMVQYAVWTTLAAVGVGANLQHY
      orf6a      AGAATILFYEDQNVVKGLEQFPAYAAFPVWADQANAMVQYAVWTTLAAVGVGANLQHY
                    100     110     120     130     140     150
10
      orf6.pep      NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
      orf6a      NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
                    160     170     180     190     200
15

```

The complete length ORF6a nucleotide sequence <SEQ ID 659> is:

```

      1  ATGACCCGTC AATCTCTGCA ACAGGCTGCC GAAAGCCGCC GTTCCATTTA
      51  TTCGTAAAT AAAAATCTGC CCGTCGGCAA AGATGAAATC GTCCAAATCG
20  101  TCGAACACGC CGTTTTCAC ACACCTTCTT CGTCAATTC CCAATCTGCC
      151  CGTGTGGTCG TGCTGTTTGG CGAAGAGCAT GATAAGGTGT GGCAATTTGT
      201  CGAAGACGCG CTGCGTGCCG TCGTGCCTGC CGACAGTTT GAACCGACCG
      251  CGCAAAATTT GAACCTGTTT AAGGCGGGTG CGGCAACTAT TTTGTTTTAT
25  301  GAAGATCAAA ATGTCGTCAA AGGTTTGAG GAGCAGTTCC CTGCTTATGC
      351  CGCCAACTTT CCCGTTTGGG CGGACCAGGC GAACGCGATG GTGCAGTATG
      401  CCGTTTGGAC GACACTTGCC GCGGTCGGCG TAGGTGCAAA CCTGCAACAT
      451  TACAATCCCT TGCCCGATGC GCGGATTGCC AAAGCGTGGA ATATCCCCGA
      501  AAAGTGGTTG TTGCGCGCAC AAATGGTTAT CGGCGGTATT GAAGGGGCGG
30  551  CAGGTGAAAA GACCTTTGAA CCAGTTGCAG AACGTTTGA AGTGTTCCGG
      601  GCATAA

```

This is predicted to encode a protein having amino acid sequence <SEQ ID 660>:

```

      1  MTRQSLQQA ESRRSIYSLN KNLPGKDEI VQIVEHAVLH TPSSFNSQSA
      51  RVVVLFGEEH DKVVQFVEDA LRAVVPADSF EPTAQKLNLF KAGAATILFY
35  101  EDQNVVKGLEQFPAYAAFPVWADQANAMVQYAVWTTLAAVGVGANLQHY
      151  YNPLPDAAIA KAWNIPENWL LRAQMVIIGI EGAAGEKTFF PVAERLKVFG
      201  A*

```

ORF6a and ORF6-1 show 100.0% identity in 131 aa overlap:

```

40  orf6a.pep      TPSSFNSQSARVVVLFGEEDKVVQFVEDALRAVVPADSFEPTAQKLNLFKAGAATILFY
      orf6-1      LRAVVPADSFEPTAQKLNLFKAGAATILFY
                        10      20      30
45  orf6a.pep      EDQNVVKGLEQFPAYAAFPVWADQANAMVQYAVWTTLAAVGVGANLQHYNPLPDAAIA
      orf6-1      EDQNVVKGLEQFPAYAAFPVWADQANAMVQYAVWTTLAAVGVGANLQHYNPLPDAAIA
                        40      50      60      70      80      90
50  orf6a.pep      KAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
      orf6-1      KAWNIPENWLLRAQMVIIGGIEGAAGEKTFFPVAERLKVFGAX
                        100     110     120     130
55

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF6 shows 95.7% identity over a 140aa overlap with a predicted ORF (ORF6ng) from *N.gonorrhoeae*:

-378-

```

      orf6.pep      KVVQFVEXPLRAVVPADSFEPTAQKLNLFK      30
                   ||||| ||||| ||||| ||||| |||||
      orf6ng      SNVSLDMSNPTVLRMGLPLYIASLRRGAIYKVVQFVEDALRAVVPADSFEPTAQKLNLFK      64
5      orf6.pep      AGAATILFYEDQNVVKGLQEQQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHY      90
                   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      orf6ng      AGAATILFYEDQNVVKGLQEQQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHY      124
10     orf6.pep      NPLPDAAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGA      140
                   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
      orf6ng      NPLPDVAIAKAWNIPENWLLRAQMVIIGGIEGAAGEKVFEPVAERLKVFGA      174

```

The complete length ORF6ng nucleotide sequence <SEQ ID 661> was identified as:

```

15      1 ATGGCCGTTG CGTCAAATGT CAGCTTGGAT ATGTCCAATC CTACGGTGT
      51 ACGCATGGGA TTACCTTAT ATATTGCGTC CCTAAGAAGG GGCGCAATAT
      101 ATAAGGTGTG GCAATTTGTC GAAGACGCGC TCGTGCCGT CGTGCCTGCC
      151 GACAGTTTGT AACCGACCGC GCAAAAATTG AAGCTGTTTA AGGCGGGCGC
      201 GGCAACCATT TTGTTTATG AAGATCAAAA TGTCGTCAAA GGTTCGAGG
      251 AGCAGTTCCT TGCTTATGCC GCCAACTTTC CCGTTTGGGC GGACCAAGCG
20     301 AACGCTATGG TACAGTATGC CGTCTGGACG ACACTTGCCG CGGTGCGTGC
      351 AGGTGCAAAT CTGCAACATT ACAACCCCTT GCCCGATGTG GCGATTGCTA
      401 AAGCGTGGAA TATCCCGAA AACTGGCTGT TCGCGCGCA AATGGTTATC
      451 GGTGGTATTG AAGGGGcggc aggtgaaaaa gtctttgaac CCGTTGCgga
      501 acgtttgAAA GTGTTGCGCG CATAA

```

25 This encodes a protein having amino acid sequence <SEQ ID 662>:

```

30      1 MAVASNVSLD MSNPTVLRMG LPLYIASLRR GAIYKVVQFV EDALRAVVA
      51 DSFEPTAQKL KLFKAGAATI LFYEDQNVVK GLQEQQPAYA ANFPVWADQA
      101 NAMVQYAVWT TAAVGVGAN LQHYNPLPDV AIAKAWNIP E NLLRAQMVI
      151 GGIEGAAGEK VFEPVAERLK VFGA*

```

ORF6ng and ORF6-1 show 96.9% identity in 131 aa overlap:

```

      orf6-1.pep      10      20      30
                   LRAVVPADSFEPTAQKLNLFKAGAATILFY
35     orf6ng      PTVLRMGLPLYIASLRRGAIYKVVQFVEDALRAVVPADSFEPTAQKLNLFKAGAATILFY
                   20      30      40      50      60      70
      orf6-1.pep      40      50      60      70      80      90
      EDQNVVKGLQEQQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHYNPLDAAIA
40     orf6ng      EDQNVVKGLQEQQFPAYAAANFPVWADQANAMVQYAVWTTTAAVGVGANLQHYNPLDVAIA
                   80      90      100      110      120      130
      orf6-1.pep      100      110      120      130
      KAWNIPENWLLRAQMVIIGGIEGAAGEKTFEPVAERLKVFGAX
45     orf6ng      KAWNIPENWLLRAQMVIIGGIEGAAGEKVFEPVAERLKVFGAX
                   140      150      160      170

```

50 It is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 79

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 663>

```

55      1 ..GGCTACAAC ACCTGTTGCG GCGCGGCAGC CGCATCGCCA ACTACCAAAT
      51 CAACGGCATC CCGTGTGCGG ACGCGCTGGC CGATACGGG CAATGCCAAC
      101 ACCGCCGCGT ATGAGCGCGT AGAAGTCGTG CGCGGCGTGG CGGGGCTGCT
      151 GGACGGCAGC GCGGAGCCTT CCGCCACCGT CAATCTGGTG CGCAAACGCC
      201 TGACCCGCAA GCCATTGTTT GAAGTCCGCG CCGAAGCGGG CAACCGCAA

```

251 CATTTCGGGC TGGACGCGGA CGTATCGGGC AGCCTGAACA CCGAAG.crc
 301 rCTGCGCgGC CGCCTGGTTT CCacCTTCGG ACgCGGCGAC TCGTGGCGGC
 351 GGCGCGAAGC CAGCCGskAT GCCGAACtCT ACgGCATTTT GGAATACGAC
 401 ATCGCACCGC AAACCCGCGT CCACGCArGC ATGGACTACC AGCAGGCGAA
 5 451 AGAAACCGCC GACGCGCGCG TCAGcTACGC CGTGTACGAC AGCCAAGGTT
 501 ATGCCACCGC CTTCGGGCCG AAAGACAACC CCGCCACAAA TTGGGCGAAC
 551 AGCCACCACC GTGCGCTCAA CCTGTTGCGC GGCATCGAAC ACCGCTTCAA
 601 CCAAGACTGG AAACtCAAAG CCGAATACGA CTAC..

This corresponds to the amino acid sequence <SEQ ID 664; ORF23>:

10 1 ..GYNLYFARGS RIANYQINGI PVADALADTG NANTAAYERV EVVRGVAGLL
 51 DGTGEPsATV NLVRKRLTRK PLFEVRAEAG NRKHFGLDAD VSGSLNtEXX
 101 LRGRlVSTFG RGDsWRRRER SRxAELYGIL EYDIAPQTRV HAXMDYQQAk
 151 ETADAPLSYA VYDSQGYATA FGPKDNpATN WANSHHRAln LFAGIEHREN
 201 QDWKLKAeyD Y..

15 Further work revealed the complete nucleotide sequence <SEQ ID 665>:

1 ATGACACGCT TCAAATATTC CCTGCTGTTT GCCGCCCTGT TGCCCGTGTA
 51 CGCGCAGGCC GATGTTTCTG TTTCAGACGA CCCCAAACCG CAGGAAAGCA
 101 CTGAATTGCC GACCATCACC GTTACCGCCG ACCGCACCGC GAGTTCCAAC
 151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCATGAC
 201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC
 251 GCGACCAAAA CATCAAACCG CTCGACCGCG CCCTGTTGCA GCGCAGCGGC
 301 ACCAGCCGCC AGATTTACGG CTCCGACCGC GCGGGCTACA ACTACCTGTT
 351 CGCGCGCGGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCGTTG
 401 CCGACGCGCT GGCCGATACG GGCAATGCCA ACACGCGCGC CTATGAGCGC
 25 451 GTAGAAGTCG TGCGCGGCGT GGCGGGGCTG CTGGACGGCA CGGGCGAGCC
 501 TTCCGCCACC GTCAATCTGG TGCGCAAACG CTGACCCGC AAGCCATTGT
 551 TTGAAGTCCG CGCCGAAGCG GGCAACCGCA AACATTTGCG GCTGGACGCG
 601 GACGTATCGG GCAGCCTGAA CACCGAAGGC ACGTGC GCGCGCTGGT
 651 TTCCACCTTC GGACGCGCGC ACTCGTGGCG GCGGCGCGAA CGCAGCCGCG
 30 701 ATGCCGAAC CTACGGCATT TTGGAATACG ACATCGCACC GCAAACCCGC
 751 GTCCACGCAG GCATGGACTA CCAGCAGGCG AAAGAAACCG CCGACGCGCC
 801 GCTCAGCTAC GCCGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC
 851 CGAAAGACAA CCCC GCCACA AATTGGGCGA ACAGCCGCCA CCGTGC GCTC
 901 AACCTGTTG CCGGCATCGA ACACCGCTTC AACCAGACT GGAACTCAA
 35 951 AGCCGAATAC GACTACACC GCAGCCGCTT CCGCCAGCCC TACGGCGTAG
 1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC
 1051 GGTATTGGC ACGCCGACCC GCGCACCAC AGCGCCAGCG TGTCATTGAT
 1101 CGGCAAAATAC CGCCTGTTG GCGCGAACA CGATTTAATC GCGGGTATCA
 1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACGCAG CATCATCCCC
 40 1201 AACGCCATTC CCAACGCCTA CGAATTTTCC CGCAGGGGTG CCTACCCGCA
 1251 GCCTGCATCG TTTGCCCAA CCATCCCGCA ATACGCCACC AGCGGCGAAA
 1301 TCGCGGGCTA TCTCGCCACC CGTTTCCGCG CCGCGACAA CCTTTCGCTG
 1351 ATTTTGGGCG GACGATACAC CCGTTACCGC ACCGGCAGCT ACGACAGCG
 1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTACC CCTACACAG
 45 1451 GCATCGTGT CGACCTGACC GGCAACCTGT CTCTTACGG CTGCTACAGC
 1501 AGCCTGTTG TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA
 1551 ACCCGTAACC GGCAACAATC TGGAAAGCGG CATCAAAGGC GAATGGCTTG
 1601 AAGGCCGTCT GAACGCATCC GCGCGCGTGT ACCGCGCCCG TAAAAAAC
 1651 CTCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC
 50 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GCGGCGCGCA
 1751 TCACGCCCGA ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCCG
 1801 GACCAAGACG GCAGCCGCCT GAACCCGAC AGCGTACCCG AACGCAGCTT
 1851 CAAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGGCTGGA
 1901 CCATCGGCGC AGGCGTGCGC TGGCAGAGCG AAACCCACAC CGACCCTGCC
 55 1951 ACGTCCGCA TCCCAACCC CGCCGCCAAA GCGCGCGCG CCGACAACAG
 2001 CCGCCAAAA GCCTACGCGC TCGCGACAT CATGGCGCGT TACCGCTTCA
 2051 ATCCGCGCGC CGAACTGTG CTGAACGTGG ACAATCTGTT CAACAAACAC
 2101 TACCGCACCC AGCCGACCG CCACAGCTAC GCGCACTGC GGACAGTGAA
 2151 CGCGCGGTTT ACCTATCGGT TTAATAA

60 This corresponds to the amino acid sequence <SEQ ID 666; ORF23-1>:

1 MTRFKYSLLF AALLPVYAQA DVSVDPPKP QESTELPTIT VTADRTASSN
 51 DGYTVSGTHT FLGLPMTLRE IPQSVSVITS QMRDONIKT LDRALLQATG
 101 TSRQIYGS DR AGYNLYFARG SRIANYQING IPVADALADT GNANTAAYER
 151 VEVVRGVAGL LDGTGEPsAT VNLVRKRLTR KPLFEVRAEA GNRKHFGLD A
 65 201 DVSGSLNTEG TLRGRlVSTF GRGDsWRRRE RSRDAELyGI LEYDIAPQTR

-380-

5
10

```

251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL
301 NLFAGIEHRRF NQDWKLKAEY DYTRSRRFRQP YGVAGVLSID HNTAATDLIP
351 GYWHADPRTH SASVSLIGKY RLFGREHDLI AGINGYKYAS NKYGERSIIP
401 NAIPNAYEFS RTGAYPOPAS FAQTIPQYGT RRQIGGYLAT RFRADNLSL
451 ILGGRYTRYR TGSYDSRTQG MTYVSANRET PYTGIVFDLT GNLSLYGSYS
501 SLFVPQSQKD EHGSYLKPVV GNNLEAGIKG EWLEGRLNAS AAVYRARKNN
551 LATAAGRDPS GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR
601 DQDGSRLNPD SVPERSEKLF TAYHFAPEAP SGWTIGAGVR WQSETHTDPA
651 TLRIPNPAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVNLFNKH
701 YRTQDRHSY GALRTVNAAF TYRFK*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the ferric-pseudobactin receptor PupB of *Pseudomonas putida* (accession number P38047)

ORF23 and PupB protein show 32% aa identity in 205aa overlap:

15
20
25

```

Orf23 6 FARGSRIANYQINGIPVADALADTGNANTAAAYERVEVVRGVAGLLDGTGEP SATVNLVRK 65
++RG I NY+++G+P + L D + + A ++RVE+VRG GL+ G G PSAT+NL+RK
PupB 215 WSRGFAIQNYEVDGVPSTRL-DNYSQSMAMFDRVEIVRGATGLISGMGNPSATINLIRK 273

Orf23 66 RLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGLRVSTFXXXXXXXXXXXXXAE 125
R T + + EAGN +G DVSG L +RGR V+ +
PupB 274 RPTAEAQASITGEAGNWDRYGTGFDVSGPLTETGNIRGRFVADYKTEKAWIDRYNQSQL 333

Orf23 126 YGILEYDIAPQTRVHAXMDYQQAKETADAPLSYAVYD--SQGYATAFGPKDNPATNWAN 183
+YGI E+D++ T + Y + D+PL + S G T N A +W+
PupB 334 MYGITEFDLSEDTLLTVGFSY--LRSDIDSPLRSGLPTRFSTGERTNLKRSLSNAAPDSY 391

Orf23 184 SHHRALNLFAGIEHRRFNQDWKLKAE 208
+ H + F IE + W K E
PupB 392 NDHEQTSFFTSIEQQLGNGWSGKIE 416

```

30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF23 shows 95.7% identity over a 211aa overlap with an ORF (ORF23a) from strain A of *N.*

meningitidis:

35
40
45
50
55
60

```

orf23.pep          GYNYLFARGSRIANYQINGIPVADALADTG
|||
orf23a      QMRDQNIKALDRALLQATGTSRQIYGS DRAGYNYLFARGSRIANYQINGIPVADALADTG
          90      100      110      120      130      140

          40      50      60      70      80      90
orf23.pep      NANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDAD
|||
orf23a      NANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRPTRKPLFEVRAEAGNRKHFGLGAD
          150     160     170     180     190     200

          100     110     120     130     140     150
orf23.pep      VSGSLNTEXXLRGLRVSTFGRGDSWRRRERSRAXAELYGILEYDIAPQTRVHAXMDYQQAK
|||
orf23a      VSGSLNAEGTLRGLRVSTFGRGDSWRQRERSRDAELYGILEYDIAPQTRVHAGMDYQQAK
          210     220     230     240     250     260

          160     170     180     190     200     210
orf23.pep      ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSHHRALNLFAGIEHRRFNQDWKLKAEYD
|||
orf23a      ETADAPLSYAVYDSQGYATAFGPKDNPATNWANSHHRALNLFAGIEHRRFNQDWKLKAEYD
          270     280     290     300     310     320

orf23.pep      Y
|
orf23a      YTRSRRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTHSASVSLIGKYRLFGRHDLIA
          330     340     350     360     370     380

```

The complete length ORF23a nucleotide sequence <SEQ ID 667> is:

1 ATGACACGCT TCAAATATTC CCTGCTGTTT GCCGCCCTGT TGCCCCGTGTA
 51 CGCGCAGGCC GATGTTTCTG TTTTCAGACGA CCCAAAACCG CAGGAAAGCA
 101 CTGAATTGCC GACCATCACC GTTACCGCCG ACCGCACCGC GAGTTCCAAC
 151 GACGGCTACA CTGTTTCCGG CACGCACACC CCGCTCGGGC TGCCCCATGAC
 201 CCTGCGCGAA ATCCCGCAGA GCGTCAGCGT CATCACATCG CAACAAATGC
 251 GCGACCAAAA CATCAAAGCG CTCGACCGCG CCCTGTTGCA GGCGACCGGC
 301 ACCAGCGGCC AGATTTACGG CTCGACCGCG GCGGGCTACA ACTACCTGTT
 351 CGCGCGCGGC AGCCGCATCG CCAACTACCA AATCAACGGC ATCCCCGTG
 401 CCGACGCGCT GGCCGATACG GGCAATGCCA ACACCGCCGC CTATGAGCGC
 451 GTAGAAGTCG TGCGCGGCGT GGCGGGGCTG CTGGACGGCA CGGGCGAGCC
 501 TTCCGCCACC GTCAATCTGG TGCGCAAACG CCCGACCCGC AAGCCATTGT
 551 TTGAAGTCCG CGCCGAAGCG GGCAACCGCA AACATTTCGG GCTGGGCGCG
 601 GACGTATCGG GCAGCCTGAA TGCCGAAGGC ACGCTGCGCG GCCGCCTGGT
 651 TTCCACCTTC GGACGCGGCG ACTCGTGCGG GCAGCGCGAA CGCAGCCGCG
 701 ATGCCGAACT CTACGGCATT TTGGAATACG ACATCGCACC GCAAACCCGC
 751 GTCCACGCAG GCATGGACTA CCAGCAGGCG AAAGAAACCG CCGACGCGCC
 801 GCTCAGCTAC GCCGTGTACG ACAGCCAAGG TTATGCCACC GCCTTCGGCC
 851 CGAAAGACAA CCCGCCACA AATTGGCGGA ACAGCCGCCA CCGTGCGCTC
 901 AACCTGTTCTG CCGGCATCGA ACACCGCTTC AACCAAGACT GGAAACTCAA
 951 AGCCGAATAC GACTACACCC GCAGCCGCTT CCGCCAGCCC TACGGCGTAG
 1001 CAGGCGTGCT TTCCATCGAC CACAACACCG CCGCCACCGA CCTGATTCCC
 1051 GGTATTGGC ACGCCGACCC GCGCACCCAC AGCGCCAGCG TGTCTAAT
 1101 CGGCAAATAC CGCTGTTCG GCCCGAACA CGATTTAATC GCGGGTATCA
 1151 ACGGTTACAA ATACGCCAGC AACAAATACG GCGAACCGAG CATCATCCCC
 1201 AACGCCATTC CCAACGCCTA CGAATTTTCC CGCACGGGTG CCTACCCGCA
 1251 GCCTGCATCG TTTGCCCAA CCATCCCGCA ATACGGCACC AGGCGGCAAA
 1301 TCGGCGGCTA TCTGCCACC CGTTTCCGCG CCGCCGACAA CCTTTCGCTG
 1351 ATACTCGGCG GCAGATACAG CCGTTACCGC ACCGGCAGCT ACGACAGCCG
 1401 CACACAAGGC ATGACCTATG TGTCCGCCAA CCGTTTACC CCCTACACAG
 1451 GCATCGTGTT CGACCTGACC GGCAACCTGT CGCTTTACGG CTCGTACAGC
 1501 AGCCTGTTCTG TCCCGCAATC GCAAAAAGAC GAACACGGCA GCTACCTGAA
 1551 ACCCGTAACC GGCAACAATC TGAAGCCGG CATCAAAGGC GAATGGCTTG
 1601 AAGGCCGTCT GAACGCATCC GCGGCCGTGT ACCGCGCCCG TAAAAACAAC
 1651 CTCGCCACCG CAGCAGGACG CGACCCGAGC GGCAACACCT ACTACCGCGC
 1701 CGCCAACCAA GCCAAAACCC ACGGCTGGGA AATCGAAGTC GGGCGCCGCA
 1751 TCACGCCCCG ATGGCAGATA CAGGCAGGTT ACAGCCAAAG CAAAACCCGC
 1801 GACCAAGACG GCAGCCGCTT GAACCCGAC AGCGTACCCG AACGCAGCTT
 1851 CAAACTCTTC ACTGCCTACC ACTTTGCCCC CGAAGCCCCC AGCGGCTGGA
 1901 CCATCGGCGC AGGCGTGCGC TGGCAGAGCG AAACCCACAC CGACCTGCCC
 1951 ACGCTCCGCA TCCCAACCC CGCCGCCAAA GCCCGCGCCG CCGACAACAG
 2001 CCGCCAAAAA GCCTACGCGC TCGCCGACAT CATGGCGCGT TACCGCTTCA
 2051 ATCCGCGCGC CGAACTGTCT CTGAACGTGG ACAATCTGTT CAACAACAC
 2101 TACCGCACCC AGCCCGACCG CCACAGCTAC GCGCACTGC GGACAGTGAA
 2151 CGCGGCGTTT ACCTATCGGT TTAAATAA

45 This encodes a protein having amino acid sequence <SEQ ID 668>:

1 MTRFKYSLLF AALLPVYAQA DVSVSDDPKP QESTELPTIT VTADRTASSN
 51 DGYTVSGTHT PLGLPMTLRE IPQSVSVITS QQMRDQNIKA LDRALLQATG
 101 TSRQIYGS DR AGYNYLFARG SRIANYQING IPVADALADT GNANTAAYER
 151 VEVVRGVAGL LDGTGEP SAT VNLVRKRPT R KPLFEVRAEA GNRKHFGLGA
 201 DVSGSLNAEG TLRGRVSTF GRGDSWRQRE RSRDAELYGI LEYDIAPQTR
 251 VHAGMDYQQA KETADAPLSY AVYDSQGYAT AFGPKDNPAT NWANSRHRAL
 301 NLFAGIEHRF NQDWKLKAEY DYTRSFRQRP YGVAGVLSID HNTAATDLIP
 351 GYWHADPRTH SASVSLIGKY RLFGRHDLI AGINGYKYAS NKYGERSIIP
 401 NAIPNAYEFS RTGAYPOPAS FAQTIPQYGT RRQIGGYLAT RFRAADNLSL
 451 ILGGYSRYR TGSYSRTQG MTYVSANRFT PYTGIVFDLT GNLSLYGSYS
 501 SLFVPSQKD EHGSYLKPV T GNNLEAGIKG EWLEGRLNAS AAVYRARKNN
 551 LATAAGRDP S GNTYYRAANQ AKTHGWEIEV GGRITPEWQI QAGYSQSKTR
 601 DDGSRNLNPD SVPERSFKLF TAYHFAPEAP SGWTIGAGVR WQSETHDPDA
 651 TLRIPNPAK ARAADNSRQK AYAVADIMAR YRFNPRAELS LNVNLFNKH
 701 YRTQPRHSY GALRTVNAAF TYRFK*

ORF23a and ORF23-1 show 99.2% identity in 725 aa overlap:

		10	20	30	40	50	60
orf23a.pep	MTRFKYSLLF	AALLPVYAQA	DVSVSDDPKP	QESTELPTIT	VTADRTASSN	DGYTVSGTHT	
65 orf23-1	MTRFKYSLLF	AALLPVYAQA	DVSVSDDPKP	QESTELPTIT	VTADRTASSN	DGYTVSGTHT	
		10	20	30	40	50	60

-382-

		70	80	90	100	110	120
	orf23a.pep	PLGLPMTLREIPQSVSVIT	SQQMRDQNIKALDRALLQATGTSRQIYGSDRAGYNYLFARG				
5	orf23-1	PLGLPMTLREIPQSVSVIT	SQQMRDQNIKTLDRAALLQATGTSRQIYGSDRAGYNYLFARG				
		70	80	90	100	110	120
	orf23a.pep	130	140	150	160	170	180
	orf23a.pep	SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRPTR					
10	orf23-1	SRIANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLDGTGEP SATVNLVRKRLTR					
		130	140	150	160	170	180
	orf23a.pep	190	200	210	220	230	240
15	orf23a.pep	KPLFEVRAEAGNRKHFG	LADVSGSLNAEGTLRGR	LVSTFGRGDSWRQRERSRDAELYGI			
	orf23-1	KPLFEVRAEAGNRKHFG	LADVSGSLNTEGTLRGR	LVSTFGRGDSWRRRERSRDAELYGI			
		190	200	210	220	230	240
	orf23a.pep	250	260	270	280	290	300
20	orf23a.pep	LEYDIAPOTRVHAGMDYQ	QAKETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRAL				
	orf23-1	LEYDIAPOTRVHAGMDYQ	QAKETADAPLSYAVYDSQGYATAFGPKDNPATNWANSRHRAL				
		250	260	270	280	290	300
25	orf23a.pep	310	320	330	340	350	360
	orf23a.pep	NLFAGIEHRFNQDWK	LKAEYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH				
	orf23-1	NLFAGIEHRFNQDWK	LKAEYDYTRSRFRQPYGVAGVLSIDHNTAATDLIPGYWHADPRTH				
30		310	320	330	340	350	360
	orf23a.pep	370	380	390	400	410	420
	orf23a.pep	SASVSLIGKYRLFGREH	DLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS				
35	orf23-1	SASVSLIGKYRLFGREH	DLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS				
		370	380	390	400	410	420
	orf23a.pep	430	440	450	460	470	480
40	orf23a.pep	FAQTIPQYGTTRRQIGG	YLATRFRAADNLSLILGGRYSR	YRTGSGYDSRTQGMTYVSANRFT			
	orf23-1	FAQTIPQYGTTRRQIGG	YLATRFRAADNLSLILGGRYTR	YRTGSGYDSRTQGMTYVSANRFT			
		430	440	450	460	470	480
45	orf23a.pep	490	500	510	520	530	540
	orf23a.pep	PYTGIVFDLTGNLSLYG	SYSSLFVPSQSKDEHGSYLKPVTGNNLEAGIKGEWLEGRNLAS				
	orf23-1	PYTGIVFDLTGNLSLYG	SYSSLFVPSQSKDEHGSYLKPVTGNNLEAGIKGEWLEGRNLAS				
		490	500	510	520	530	540
50	orf23a.pep	550	560	570	580	590	600
	orf23a.pep	AAVYRARKNNLATAAGR	DPGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR				
	orf23-1	AAVYRARKNNLATAAGR	DPGNTYYRAANQAKTHGWEIEVGGRITPEWQIQAGYSQSKTR				
		550	560	570	580	590	600
55	orf23a.pep	610	620	630	640	650	660
	orf23a.pep	DQDGSRLNPDSVPERS	FKLFTAYHFAPEAPSGWTIGAGVRWQSEHTDPATLRIPNPAAK				
	orf23-1	DQDGSRLNPDSVPERS	FKLFTAYHFAPEAPSGWTIGAGVRWQSEHTDPATLRIPNPAAK				
60		610	620	630	640	650	660
	orf23a.pep	670	680	690	700	710	720
	orf23a.pep	ARAADNSRQKAYAVAD	IMARYRFNPRAELSLNVDNLFNKHYRTQDRHSYGALRTVNAAF				
	orf23-1	ARAADNSRQKAYAVAD	IMARYRFNPRAELSLNVDNLFNKHYRTQDRHSYGALRTVNAAF				
65		670	680	690	700	710	720
	orf23a.pep	TYRFXK					
70	orf23-1	TYRFXK					

Homology with a predicted ORF from *N.gonorrhoeae*

ORF23 shows 93.4% identity over a 211aa overlap with a predicted ORF (ORF23.ng) from *N.gonorrhoeae*:

5	orf23.pep	GYNYLFARGSR IANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLLD	51
	orf23ng	SAVDACRIPGYNLYFARGSR IANYQINGIPVADALADTGNANTAAYERVEVVRGVAGLPD	60
10	orf23.pep	GTGEPSATVNLVRKRLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEXXLRGLRVSTFGR	111
	orf23ng	GTGEPSATVNLVRKHPTRKPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGLRVSTFGR	120
15	orf23.pep	GDSWRRRERSRXAELYGILEYDIAPQTRVHAXMDYQQAETADAPLSYAVYDSQGYATAF	171
	orf23ng	GDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQQAETADAPLSYAVYDSQGYATAF	180
	orf23.pep	GPKDNPATNWSNSRRLNLFAGIEHRFNQDWKLKAEYDY	211
	orf23ng	GPKDNPATNWSNSRRLNLFAGIEHRFNQDWKLKAEYDYTRSRFRQPYGVAGVLSIDHS	240

The ORF23ng nucleotide sequence <SEQ ID 669> is predicted to encode a protein comprising

20 amino acid sequence <SEQ ID 670>:

	1	SAVDACRIPG	YNYLFARGSR	IANYQINGIP	VADALADTGN	ANTAAYERVE
	51	VVRGVAGLPD	GTGEPSATVN	LVRKHPTRKP	LFEVRAEAGN	RKHFGLGADV
	101	SGSLNAEGLT	RGLRVSTFGR	GDSWRQLERS	RDAELYGILE	YDIAPQTRVH
	151	AGMDYQQAKE	TADAPLSYAV	YDSQGYATAF	GPKDNPATNW	SNSRRLN
25	201	FAGIEHRFNQ	DWKLKAEYDY	TRSRFRQPYG	VAGVLSIDHS	TAATDLIPGY
	251	WHADPRTHSA	SMSLTGKYRL	FGREHDLIAG	INGYKYASNK	YGERSIIPNA
	301	IPNAYEFSRT	GAYPOPSFA	QTIPOYDTRR	QIGGYLATRF	RAADNLSLIL
	351	GGYRSRYRAG	SYNSRTQGMT	YVSANRFTPY	TGIVFDLTGN	LSLYGSYSSL
30	401	FVPQLQKDEH	GSYLKPVGTN	NLEADIKGEW	LEGRLNASAA	VYRARKNNLA
	451	TAAGRDQSGN	TYYRAANQAK	THGWEIEVGG	RITPEWQIQA	GYSQSKPRDQ
	501	DGSRLNPDSV	PERSFKLETA	YHLAPEAPSG	RTIGAGVRRQ	GETHTDPAAL
	551	RIPNPAAKAR	AVANSRQKAY	AVADIMARYR	FNPRTLSLN	VDNLFNKHYR
	601	TQPDHRSYGA	LRTVNAAFTY	RFK*		

Further work revealed the complete nucleotide sequence <SEQ ID 671>:

35	1	ATGACACGCT	TCAAATACTC	CCTGCTTTTT	GCCGCCCTGC	TACCCGTGTA
	51	CGCGCAGGCC	GATGTTTCTG	TTTCAGACGA	CCCCAAACCG	CAGGAAAGCA
	101	CCGAATTGCC	GACCATCACC	GTTACCGCCG	ACCGCACCAG	GAGTTCCAAC
	151	GACGGCTACA	CCGTTTCCGG	CACGCACACC	CCGTTCCGGC	TGCCCATGAC
40	201	CCTGCGCGAA	ATCCCGCAGA	GCGTCAGCGT	CATCACATCG	CAACAAATGC
	251	GCGACCAAAA	CATCAAAACG	CTCGACCCGC	CCCTGTTGCA	GCGGACCCGC
	301	ACCAGCCGCC	AGATTACGG	CTCCGACCGC	GCGGGCTACA	ACTACCTGTT
	351	CGCGCGCGGC	AGCCGCATCG	CCAATACCA	AATCAACGGC	ATCCCCGTTG
	401	CCGACGCGCT	GGCCGATACG	GGCAATGCCA	ACACCGCCGC	CTATGAGCGC
	451	GTAGAAGTCG	TGCGCGCGCT	GGCGGGGCTG	CCGGACGGCA	CGGGCGAGCC
45	501	TTCTGCCACC	GTCAATCTGG	TACGCAAACA	CCCGACCCGC	AAGCCATTGT
	551	TTGAAGTCCG	CGCCGAAGCC	GGCAACCGCA	AACATTTCCG	GCTGGGCGCG
	601	GACGTATCGG	GCAGCCTGAA	CGCCGAAGGC	ACGCTGCGCG	GCCGCCTGGT
	651	TTCCACCTTC	GGACGCGGCG	ACTCGTGCGG	GCAGCTCGAA	CGCAGCCGCG
	701	ATGCCGAATC	CTACGGCATT	TTGGAATACG	ACATCGCACC	GCAAACCCGC
	751	GTCCACGCAG	GCATGGACTA	CCAGCAGGCG	AAAGAAACCG	CAGACGCGCC
50	801	GCTCAGCTAC	GCCGTGTACG	ACAGCCAAGG	TTATGCCACC	GCCTTCGGCC
	851	CAAAAGACAA	CCCCGCCACA	AATTGGTCGA	ACAGCCGCAA	CCGTGCGCTC
	901	AACCTGTTCT	CCGGCATAGA	ACACCGCTTC	AACCAAGACT	GGAAACTCAA
	951	AGCCGAATAC	GACTACACCC	GTAGCCGCTT	CCGCCAGCCC	TACGGTGTGG
55	1001	CAGGCGTACT	TTCCATCGAC	CACAGCACTG	CCGCCACCGA	CCTGATTCCC
	1051	GGTTATTGGC	ACGCGgatacc	GCGCACCCAC	AGCGCCAGCA	TGTCATTGAC
	1101	CGGCAAATAC	GgcctGTTCT	GCCGCGAGCA	CGATTAAATC	GCGGGTATCA
	1151	ACGGCTACAA	ATACGCCAGC	AACAAATACG	GCGAACGCAG	CATCATTTCC
	1201	AACGCCATTTC	CCAACGGCTA	CGAATTTTCC	CGCACGGGGC	CCTATCCGCA
	1251	GCCATCATCG	TTTGCCCAAA	CCATCCCGCA	ATACGACACC	AGGCGGCAAA
60	1301	TCGGCGGCTA	TCTCGCCACC	CGTTTCCGCG	CCGCCGACAA	CCTTTGCTGT
	1351	ATACTCGGCG	GCAGATACAG	CCGCTACCCG	GCAGGCAGCT	ACAACAGCCG

1401	CACACAAGGC	ATGACCTATG	TGTCCGCAA	CGGTTTCACC	CCCTACACAG
1451	GCATCGTGTT	CGATCTGACC	GGCAACCTGT	CGCTTTACGG	CTCGTACAGC
1501	AGCCTGTTTCG	TCCCGCAATT	GCAAAAAGAC	GAACACGGCA	GCTACCTGAA
1551	ACCGTAACC	GGCAACAATC	TGGAAGCCGA	CATCAAAGGC	GAATGGCTTG
1601	AAGGGCGTCT	GAACGCATCC	GCCGCGGTGT	ACCGCGCCCG	TAAAAACAAC
1651	CTCGCCACCG	CAGCAGGACG	CGACCAGAGC	GGCAACACCT	ACTATCGCGC
1701	CGCCAACCAA	GCCAAAACCC	ACGGCTGGGA	AATCGAAGTC	GGCGGCCGCA
1751	TCACGCCCCA	ATGGCAGATA	CAGGCAGGCT	ACAGCCAAAG	CAAACCCCGC
1801	GACCAAGACG	GCAGCCGCCT	GAACCCCGAC	AGCGTACCCG	AACGCAGCTT
1851	CAAACCTCTC	ACCGCCTACC	ACTTAGCCCC	CGAAGCCCCC	AGCGGCCGGA
1901	CCATcggTGC	GGGTGTGCGC	CGGCAGGGCG	AAACCCACAC	CGACCCAGCC
1951	CGGCTCCGCA	TCCCCAACCC	CGCCGCCAAA	GCCCGCGCGG	TCGCCAACAG
2001	CCGCCAGAAA	GCCTACGCCG	TCGCCGACAT	CATGGCGCGT	TACCGCTTCA
2051	ATCCGCGCAC	CGAAGTGTGC	CTGAACGTGG	ACAACCTGTT	CAACAAACAC
2101	TACCGCACCC	AGCCCGACCG	CCACAGCTAC	GGCGCACTGC	GGACAGTGAA
2151	CGCGGCGTTT	ACCTATCGGT	TTAAATAA		

This corresponds to the amino acid sequence <SEQ ID 672; ORF23ng-1>:

1	MTRFKYSLLF	AALLPVYAQA	DVSVSDDPKP	QESTELPTIT	VTADRTASSN
51	DGYTVSGTHT	PFGLPMTLRE	IPQSVSVITS	QQMRDQNIKT	LDRALLQATG
101	TSRQIYGS DR	AGYNYLFARG	SRIANYQING	IPVADALADT	GNANTAAYER
151	VEVVRGVAGL	PDGTGEP SAT	VNLVRKHPT R	KPLFEVRAEA	GNRKHFGLGA
201	DVSGSLNAEG	TLRGRLVST F	GRGDSWRQLE	RSRDAELYGI	LEYDIA PQTR
251	VHAGMDYQQA	KETADAPLS Y	AVYDSQGYAT	AFGPKDNPAT	NWSNSRNRAL
301	NLFAGIEHRF	NQDWKLKAEY	DYTRSFRQP Y	YGAVGVLSID	HSTAATDLIP
351	GYWHADPRTH	SASMSLTGKY	RLFGREHDLI	AGINGYKYAS	NKYGERSIIP
401	NAIPNAYEFS	RTGAYQPSS	FAQTIPOYDT	RRQIGGYLAT	RFRAADNLSL
451	ILGGYRSYR	AGSYNSRTQG	MTYVSANRFT	PYTGIVFDLT	GNLSLYGSYS
501	SLFVPOLQKD	EHGSYLKPV T	GNNLEADIKG	EWLEGRNLAS	AAVYRARKNN
551	LATAAGRDQS	GNTYYRAANQ	AKTHGWEIEV	GGRITPEWQI	QAGYSQSKPR
601	DQDGSRLNPD	SVPERSFKLF	TAYHLAPEAP	SGRTIGAGVR	RQGETHTDPA
651	ALRIPNPAAK	ARAVANSRQK	AYAVADIMAR	YRFNPRTELS	LNVDNLFNKH
701	YRTQPDHSY	GALRTVNAAF	TYRFK*		

ORF23ng-1 and ORF23-1 show 95.9% identity in 725 aa overlap:

35	orf23-1.pep	10	20	30	40	50	60
	orf23ng-1	10	20	30	40	50	60
40	orf23-1.pep	70	80	90	100	110	120
	orf23ng-1	70	80	90	100	110	120
45	orf23-1.pep	130	140	150	160	170	180
	orf23ng-1	130	140	150	160	170	180
50	orf23-1.pep	190	200	210	220	230	240
	orf23ng-1	190	200	210	220	230	240
55	orf23-1.pep	250	260	270	280	290	300
	orf23ng-1	250	260	270	280	290	300
60	orf23-1.pep	310	320	330	340	350	360
	orf23ng-1	310	320	330	340	350	360

-385-

		310	320	330	340	350	360
5	orf23-1.pep	370	380	390	400	410	420
	orf23ng-1	SASVSLIGKYRLFGREHDLIAGINGYKYASNKYGERSIIPNAIPNAYEFSRTGAYPQPAS					
		370	380	390	400	410	420
10	orf23-1.pep	430	440	450	460	470	480
	orf23ng-1	FAQTIPQYDTRRQIGGYLATRFRAADNLSLILGGRYTRYRTGSYDSRTQGMTYVSANRFT					
		430	440	450	460	470	480
15	orf23-1.pep	490	500	510	520	530	540
	orf23ng-1	PYTGIVFDLTGNLSLYGSYSSLFVPQSQKDEHGSYLKPVTGNNLEAGIKGEWLEGRNAS					
		490	500	510	520	530	540
20	orf23-1.pep	550	560	570	580	590	600
	orf23ng-1	AAVYRARKNNLATAAGRDPSGNTYRAANQAKTHGWEIEVGGRIITPEWQIQAGYSQSKTR					
25	orf23-1.pep	550	560	570	580	590	600
	orf23ng-1	AAVYRARKNNLATAAGRDQSGNTYRAANQAKTHGWEIEVGGRIITPEWQIQAGYSQSKPR					
		550	560	570	580	590	600
30	orf23-1.pep	610	620	630	640	650	660
	orf23ng-1	DQDGSRLNPDSVPERSFKLFTAYHFAPEAPSGWTIGAGVVRQSEHTDPAALRIPNPAK					
		610	620	630	640	650	660
35	orf23-1.pep	670	680	690	700	710	720
	orf23ng-1	ARAADNSRQKAYAVADIMARYRFNPRAELSLNVDNLFNKHYRTQPDHRSYGALRTVNAAF					
		670	680	690	700	710	720
40	orf23-1.pep	TYRFXK					
	orf23ng-1	TYRFXK					

In addition, ORF23ng-1 shows significant homology with an OMP from *E.coli*:

45	sp P16869 FHUE_ECOLI OUTER-MEMBRANE RECEPTOR FOR FE(III)-COPROGEN, FE(III)-FERRIOXAMINE B AND FE(III)-RHODOTRULIC ACID PRECURSOR >gi 1651542 gnl PID d1015403 (D90745) Outer membrane protein FhuE precursor [Escherichia coli]
50	>gi 1651545 gnl PID d1015405 (D90746) Outer membrane protein FhuE precursor [Escherichia coli] >gi 1787344 (AE000210) outer-membrane receptor for Fe(III)-coprogen, Fe(III)-ferrioxamine B and Fe(III)-rhodotrulic acid precursor [Escherichia coli] Length = 729 Score = 332 bits (843), Expect = 3e-90 Identities = 228/717 (31%), Positives = 350/717 (48%), Gaps = 60/717 (8%)
55	Query: 38 TITVTADRTASSN--DGYTVSGTHTPFGLPMTLREIPQSVSVITSQQMRDONIKTLDRAL 95 T+ V TA + + Y+V+ T + MT R+IPQSV++++ Q+M DQ ++TL + Sbjct: 43 TVIVEGSATAPDDGENDYSVTSTSAQTKMQMTQRDIPQSVTIVSQORMEDQQLQTLGEVM 102
60	Query: 96 LQATGTSRQIYGSDRAGYNLYFARGSRANYQINGIP-----VADALADTGNANTAA 147 G S+ SDRA Y ++RG +I NY ++GIP + DAL+D A Sbjct: 103 ENTLGISKSQADSRLALY---YSRGGQIDNYMVDGIPTYFESRWNLGDALSDM-----AL 154
65	Query: 148 YERVEVVRGVAGLPDGTGEPSTVNLVRKHPTKPLF-EVRAEAGNRKHFLGADVSGSL 206 +ERVEVVRG GL GTG PSA +N+VRKH T + +V AE G+ AD+ L Sbjct: 155 FERVEVVRGATGLMTGTGNPSAAINMVRKHATSREFKGDVSAEYGSWNKERYVADLQSPL 214
70	Query: 207 NAEGTLRGLRVSTFGRGDSWRQLERSRDAELYGILEYDIAPQTRVHAGMDYQAKETADA 266 +G +R R+V + DSW S GI++ D+ T + AG +YQ+ + Sbjct: 215 TEDGKIRARIVGGYQNNDSWLDRYNSEKTFEFGIVDADLGDLTLSAGYEYQRIDVNSPT 274
	Query: 267 PLSYAVYDSQGYATAFGPKDNPATNWSNSRNALNLFAGIEHRFNQDWKLKAEYDYTRSR 326

-386-

+++ G + ++ + A +W+ + +F ++ +F W+ ++
 Sbjct: 275 WGGLPRWNTDGSSNSYDRARSTAPDWAYNDKEINKVFMTLKQQFADTWQATLNATHSEVE 334
 Query: 327 F--RQPYGVAGVLSIDHSTAA--TDLIPGY-----WHADPRTHSA-SMSLTGKYRLFG 374
 F + Y A V D ++ PG+ W++ R A + G Y LFG
 Sbjct: 335 FDSKMMYVDAYVNKADGMLVGPYSNYGPGFDYVGGTGWNSGKRKVDALDLFADGSYELFG 394
 Query: 375 REHDLIAGINGYKYASNKYGER--SIIPNAIPNAYEFSRTGAYPQPSSFAQTIPQYDTRR 432
 R+H+L+ G Y +N+Y +I P+ I + Y F+ G +PQ Q++ Q DT
 Sbjct: 395 RQHNLMFQ-GSYSKQNNRYFSSWANIFPDEIGSFYFN--GNFPQTDWSPQSLAQDDTTH 451
 Query: 433 QIGGYLATRFRAADNLSLILGGYRSYRAGSYNSRTQGMTY-VSANRETPTYGIVFDXXX 491
 Y ATR AD L LILG RY+ +R + +TY + N TPY G+VFD
 Sbjct: 452 MKSLYAATRVTLADPLHLILGARYTNWRVDT-----LTYSMEKNHTTPYAGLVFDIND 504
 Query: 492 XXXXXXXXXXXXFPQLQKDEHGSYLKPVGTNNLEADIKGEWLEGRNLASAAVYRARKNNL 551
 F PQ +D G YL P+TGNN E +K +W+ RL + A++R ++N+
 Sbjct: 505 NWSTYASYTSIFQPQNDRDSSGKYLAPITGNNYELGLKSDWMNSRLTTTLAIFRIEQDNV 564
 Query: 552 ATAAGR---DQSGNTYYRAANQAKTHGWEIEVGGRTPEWQIQAGYSQSKPRDQDGSRLN 608
 A + G +G T Y+A + + G E E+ G IT WQ+ G ++ D +G+ +N
 Sbjct: 565 AQSTGTPIPGSNGETAYKAVDGTVSKGVEFELNGAITDNWQLTFGATRYIAEDNEGNAVN 624
 Query: 609 PDSVPERSFKLFTAYHLAEPASGRITIGAVRRQGETHTDPAALRIPNPAAKARAVANSR 668
 P ++P + K+FT+Y L P P T+G GV Q +TD P RA
 Sbjct: 625 P-NLPRTTVKMFTSYRL-PVME-LTVGGGVNWQNRVYTDTV-----TPYGTFR-----E 672
 Query: 669 QKAYAVADIMARYRFNPELSELNVNDFLNKHYRTQPDRLH-SYGALRTVNAAFTYRF 724
 Q +YA+ D+ RY+ L NV+NLF+K Y T + YG R + TY+F
 Sbjct: 673 QGSYALVDLFTRYQVTKNFSLQGNVNNLFDKTYDTNVEGSIVYGTPRNFSITGTYQF 729

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF23-1 (77.5kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 15A shows the results of affinity purification of the His-fusion protein, and Figure 15B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 15C) and for ELISA (positive result). These experiments confirm that ORF23-1 is a surface-exposed protein, and that it is a useful immunogen.

40 Example 80

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 673>:

1 ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
 51 GGCAATGATG CCGGAAATGG TGTGCGCGGG CGTGTGCGCG GGAACGGCAA
 101 TCATATCCAA GCCGACCGAA CAAACGGCGG TCATGGCTTC GAGTTTGTC
 151 AGCGTCAGcA CGCCTGCTTC GGCGgcGgCa ATCATACCTT CGTCTTCGGA
 201 AACGGGGATA AACGcGCCAC TCAAACCCCG GACCGCGCTG GAAGCCATCA
 251 TGCCGCCTTT TTTCACGGCA TCGTTCAGCA ATGCCAAAGC TGCTGTTGTG
 301 CCGTGCGTAC CGCAGACGCT CAAGCCCAT TnTTCAAGAA TGCGTGCCAC
 351 TnAGTCGCGG ACGGGG..

50 This corresponds to the amino acid sequence <SEQ ID 674; ORF24>:

1 MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISKPTQ QTAVMASSLS
 51 SVSTPASAAA IIPSSSETGI NAPLKPPTAL EAIMPPFFTA SFSNAKAADV
 101 PCVPQTLKPI XSRMRATXSP TG..

Further work revealed the complete nucleotide sequence <SEQ ID 675>:

```

1  ATGCGCACGG  CAGTGGTTTT  GCTGTTGATC  ATGCCGATGG  CGGCTTCGTC
51  GGCAATGATG  CCGGAAATGG  TGTGCGCGGG  CGTGTGCGCG  GGAACGGCAA
101 TCATATCCAA  GCCGACCGAA  CAAACGGCGG  TCATGGCTTC  GAGTTTGTCC
151 AGCGTCAGCA  CGCCTGCTTC  GCGGCGGGCA  ATCATACCTT  CGTCTTCGGA
201 AACGGGGATA  AACGCGCCAC  TCAAACCCCC  GACCGCGCTG  GAAGCCATCA
251 TGCCGCCTTT  TTTCACGGCA  TCGTTCAGCA  ATGCCAAAGC  TGCTGTGTG
301 CCGTGCGTAC  CGCAGACGCT  CAAGCCCAT  TCTTCAAGAA  TGCCTGCCAC
351 TGAGTCGCGG  ACGGCGGGGG  TCGGCGCCAG  CGACAAGTCG  AGAATACCAA
401 ACGGGATATT  CAGCATTTTT  GAGGCTTCGC  GGCCGATGAG  TTCGCCACG
451 CGGGTAATTT  TGAAAGCAGT  TTTCTTCACT  ACTTCCGCAA  CTTCGGTCAA
501 TGTGCTTGCA  TCTGAATTTT  CCAACGCGGC  TTTTACGACA  CCTGGGCCGG
551 ATACGCCGAC  ATTGATAACG  GCATCCGCTT  CGCCCGAACC  ATGAAACGCG
601 CCCGCCATAA  ACGGGTTGTC  TTCCACCGCG  TTGCAGAACA  CGACAATTTT
651 AGCGCAGCCG  AAACCTTCGG  GCGTGATTTC  CGCCGTGCGT  TTGACGGTTT
701 CGCCCGCCAG  CTTGACCGCA  TCCATATTGA  TACCGGCACG  CGTACTGCCG
751 ATATTGATGG  AGCTGCACAC  AATATCGGTA  GTCTTCATCG  CTTGCGGAAT
801 GGAGCGGATT  AACACCTCAT  CCGAAGGCGA  CATCCCTTTT  TGCACCAACG
851 CGGAAAACCC  GCCGATAAAA  GACACACCGA  TGGCTTTGGC  AGCTTTATCC
901 AAAGTTTGCG  CCACGCTGAC  GTAA

```

This corresponds to the amino acid sequence <SEQ ID 676; ORF24-1>:

```

1  MRTAVVLLLI  MPMAASSAMM  PEMVCAGVSP  GTAIISKPTE  QTAVMASSLS
51  SVSTPASAAA  IIPSSSETGI  NAPLKPPTAL  EAIMPPFFTA  SFSNAKAADV
101 PCVPQTLKPI  SSRMRATESP  TAGVGASDKS  RIPNGIFSIF  EASRPMSSPT
151 RVILKAVFFT  TSATSVNVVA  SEFSNAAFTT  PGPDTPTLIT  ASASPEP*NA
201 PAINGLSSTA  LQNTTILAQP  KPSGVISAVR  LTVSPASLTA  SILIPARVLE
251 ILMELHTISV  VFIASGMERI  NTSSEGDIPF  CTNAEKPPIK  DTPMALAALS
301 KVCATLT*

```

Computer analysis of this amino acid sequence gave the following results:

30 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF24 shows 96.4% identity over a 307 aa overlap with an ORF (ORF24a) from strain A of *N. meningitidis*:

```

35  orf24a.pep  10      20      30      40      50      60
    MRTAVVLLIIMPMAASSAMMPMVCAVSPGTAIISKXPTEQTAVIASSLSNVSTPASAAA
    orf24      10      20      30      40      50      60
    MRTAVVLLIIMPMAASSAMMPMVCAVSPGTAIISKXPTEQTAVMASSLSNVSTPASAAA

40  orf24a.pep  70      80      90      100     110     120
    IIPSSSXTGINAPLKPPTALEAIMPFFFTASFSNAKAADVPCVPQTLKPISSMRATESP
    orf24      70      80      90      100     110     120
    IIPSSSXTGINAPLKPPTALEAIMPFFFTASFSNAKAADVPCVPQTLKPISSMRATESP

45  orf24a.pep  130     140     150     160     170     180
    TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
    orf24      130     140     150     160     170     180
    TAGVGASDKSRIPNGIFSIFEASRPMSSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT

50  orf24a.pep  190     200     210     220     230     240
    PGPDTPTLITASASPEPXNAPAIKGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
    orf24      190     200     210     220     230     240
    PGPDTPTLITASASPEPXNAPAIKGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA

55  orf24a.pep  250     260     270     280     290     300
    SILIPARVLPILMELHTISVVFIAAGMERXNTSSEGDIPFCTSAEKPPIKDTPMALAALS
    orf24      250     260     270     280     290     300
    SILIPARVLPILMELHTISVVFIAAGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS

```

```

orf24a.pep  KVCATLTX
|||||||
orf24       KVCATLTX

```

5 The complete length ORF24a nucleotide sequence <SEQ ID 677> is:

```

      1 ATGCGCACGG CAGTGGTTTT GCTGTTGATC ATGCCGATGG CGGCTTCGTC
     51 GGCAATGATG CCGGAAATGG TGTGCGCGGG TGTGTCGCCG GGAACGGCAA
    101 TCATATCCAA NCCGACCGAA CAAACGGCGG TCATCGCTTC GAGTTTATCC
    151 AACGTCAGCA CGCCTGCTTC GGCGGCGGCA ATCATACCTT CGTCTTCGGA
    201 NACGGGGATA AACGCGCCAC TCAAACCGCC AACC CGCTC GAAGCCATCA
    251 TGCCGCCCTT TTTCACGGCA TCGTTCAGCA ATGCCAAGC TGCTGTTGTG
    301 CCGTGCCTAC CGCAGACGCT CAAACCCATT TCTTCAAGAA TGCGCGCCAC
    351 CGAGTCGCCG ACGGCAGGGG TCGGTGCCAG CGACAAGTCG AGAATACCAA
    401 ACGGGATATT CAGCATTTTT GAGGCTTCGC GGCCGATGAG TTCGCCACG
    451 CGGGTAATTT TGAAGGCGGT TTTCTTACA ACTTCGGCAA CTTCGGTCAA
    501 TGTCGTTGCA TCCGAATTTT CCAACGCGGC TTTTACGACA CCCGGGCCGG
    551 ATACGCCGAC ATTAATCACA GCATCCGCTT CGCCTGAGCC GTGAAACGCG
    601 CCCGCCATAN ACGGGTTGTC TTCCNCCGCG TTGCAGAAAC CGACGATTTT
    651 GGCGCAGCCG AAACCTTTCTA GTGTGATTT ANCCGTGCGT TTGATGGTTT
    701 CGCCCGCCAG TCTGACCGCG TCCATATTGA TACCGGCGCG CGTACTGCCG
    751 ATATTGATGG AGCTGCACAC GATATCAGTA GTCTTCATCG CTTCGGGAAT
    801 GGAACGATN AACACCTCGT CAGAAGGCGA CATACCTTTT TGCACGACG
    851 CGGAAAAGCC GCCAATAAAA GACACGCCGA TGGCTTTGGC AGCCTTATCC
    901 AAAGTTTGGC CCACGCTGAC GTAA

```

25 This encodes a protein having amino acid sequence <SEQ ID 678>:

```

      1 MRTAVVLLLI MPMAASSAMM PEMVCAGVSP GTAIISXPTE QTAVIASSLS
     51 NVSTPASAAA IIPSSSXTGI NAPLKPPTAL EAIMPPFFTA SFSNAKAADV
    101 PCVPQTLKPI SSRMRATESP TAGVGASDKS RIPNGIFSIF EASRPMSPT
    151 RVILKAVFFT TSATSVNVVA SEFSNAAFTT PGPDTPTLIT ASASPEP*NA
    201 PAIXGLSSXA LQNTTILAQP KPSSVISXVR LMVSPASLTA SILIPARVLP
    251 ILMELHTISV VFIASGMERX NTSSEGDIPF CTSAEKPPIK DTPMALAALS
    301 KVCATLT*

```

It should be noted that this protein includes a stop codon at position 198.

ORF24a and ORF24-1 show 96.4% identity in 307 aa overlap:

```

35      10      20      30      40      50      60
orf24a.pep  MRTAVVLLLIMPMAASSAMMPMV CAGVSPGTAIISXPTEQTAVIASSLSNVSTPASAAA
          |||
orf24-1      MRTAVVLLLIMPMAASSAMMPMV CAGVSPGTAIISKPTEQTAVMASSLSNVSTPASAAA
          |||
      10      20      30      40      50      60
40      70      80      90     100     110     120
orf24a.pep  IIPSSSXTGINAPLKPPTALEAIMPPFFTASFSNAKAADVPCVPQTLKPISSRMRATESP
          |||
orf24-1      IIPSSSETGINAPLKPPTALEAIMPPFFTASFSNAKAADVPCVPQTLKPISSRMRATESP
          |||
45      70      80      90     100     110     120
      130     140     150     160     170     180
orf24a.pep  TAGVGASDKSRIPNGIFSIFEASRPMSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
          |||
50      130     140     150     160     170     180
orf24-1      TAGVGASDKSRIPNGIFSIFEASRPMSPTRVILKAVFFTTSATSVNVVASEFSNAAFTT
          |||
      190     200     210     220     230     240
      190     200     210     220     230     240
55      190     200     210     220     230     240
orf24a.pep  PGPDTPTLITASASPEPXNAPAI XGLSSXALQNTTILAQPKPSSVISXVRLMVSPASLTA
          |||
orf24-1      PGPDTPTLITASASPEPXNAPAI NGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA
          |||
      250     260     270     280     290     300
60      250     260     270     280     290     300
orf24a.pep  SILIPARVLPILMELHTISVVFIASGMERXNTSSEGDIPFCTSAEKPPIKDTPMALAALS
          |||
orf24-1      SILIPARVLPILMELHTISVVFIASGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS
          |||

```

```

      orf24a.pep      KVCATLTX
      |||||
5      orf24-1      KVCATLTX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF24 shows 96.7% identity over a 121 aa overlap with a predicted ORF (ORF24ng) from *N.gonorrhoeae*:

```

10      orf24.pep      MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIISKPTQAVMASSLSVSTPASAAA      60
      orf24ng      MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIMSKPTQAVMASSLSVSTPASAAA      60

15      orf24.pep      IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVVPCVPQTLKPIIXSRMRATXSP      120
      orf24ng      IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVVPCVPQTLKPISSRMRAATESP      120

      orf24.pep      TG      122
      |:
20      orf24ng      TAGVGASDKSRMPNGIFSIFEASRPMSSPTRVILKAVFFTTSSATSVRLTASEFSSAALT      180

```

The complete length ORF24ng nucleotide sequence <SEQ ID 679> is:

```

      1  ATGCGCACGG  CGGTGGTTTT  GCTGTTGATC  ATGCCGATGG  CGGCTTCGTC
51  GCGGATGATG  CCGGAAATGG  TGTGCGCGGG  CGTGTGCGCG  GGAACGGCAA
101 TCATGTCCAA  ACCAACGGAG  CAGACGGCGG  TCATGGCTTC  GAGTTTGTCC
25  151 AGCGTCAACA  CGCCTGCCTC  GCGGGCGGCA  ATCATACCTT  CGTCTTCGGA
201 AACGGGGATA  AACCGCGCCG  TCAAACCGCC  GACCGCGCTG  GAAGCCATCA
251 TGCCGCCCTT  TTTCACGGCA  TCGTTCAGCA  ATGCCAAAGC  TGCTGTTGTG
301 CCGTGCGTAC  CGCAGACGCT  CAAGCCCAT  TCTTCAAGAA  TGC GCGCCAC
351 CGAGTCGCGG  ACGGCGGGGG  TCGGTGCCAG  CGACAAATCG  AGAATGCCGA
10  401 ACGGGATATT  CAGCATTTTT  GAGGCTTCGC  GACCGATGAG  TTCGCCCACG
451 CGGGTGATTT  TGAAAGCGGT  TTTCTTACG  ACTTCGGCGA  CCTCGGTCAG
501 GCTGACCGCG  TCCGAATTTT  CCAGCGCGGC  TTTGACCACG  CCTGGACCGG
551 ATACGCCGAC  ATTAATCACA  GCATCCGCTT  CGCCCGAGCC  GTGGAACGCA
601 CCCGCCATAA  ACGGATTGTC  TTCCACCGCG  TTGCAGAACA  CGACGATTTT
35  651 GCGCGAGCCG  AAACCTTCGG  GTGTGATTTC  AGCCGTGCGT  TTGATGGTTT
701 CGCCTGCCAG  CTTGACCGCA  TCCATATTGA  TACCGGCACG  CGTGCTGCCG
751 ATATTGATGG  AGCTGCACAC  GATATCGGTA  GTTTTCATCG  CTTGCGGAAC
801 GGAACGGATC  AACACCTCAT  CCGAAGGCGA  CATACCTTTT  TGCACCGCG
40  851 CGGAAAAGCC  GCCGATAAAG  GACACGCCGA  TGGCTTTGGC  TGCCTTGTC
901 AAAGTCTGCG  CCACGCTGAC  ATAA

```

This encodes a protein having amino acid sequence <SEQ ID 680>:

```

      1  MRTAVVLLLI  MPMAASSAMM  PEMVCAGVSP  GTAIMSKPTE  QTAVMASSLS
51  SVNTPASAAA  IIPSSSETGI  NAPLKPPTAL  EAIMPFFFTA  SFSNAKAAVV
101 PCVPQTLKPI  SSRMRATESP  TAGVGASDKS  RMPNGIFSIF  EASRPMSSPT
45  151 RVILKAVFFT  TSATSVRLTA  SEFSSAALT  PGPDPTLIT  ASASPEPWA
201 PAINGLSSTA  LQNTTILAQP  KPSGVISAVR  LMVSPASLTA  SILIPARVLP
251 ILMELHTISV  VFIASGTERI  NTSSEGDIPF  CTSAEKPPIK  DTPMALAALS
301 KVCATLT*

```

ORF24ng and ORF24-1 show 96.1% identity in 307 aa overlap:

```

50      orf24-1.pep      10      20      30      40      50      60
      orf24ng      MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIISKPTQAVMASSLSVSTPASAAA
      |||||
      orf24ng      MRTAVVLLLIMPMAASSAMPEMVCAGVSPGTAIMSKPTQAVMASSLSVSTPASAAA

55      orf24-1.pep      70      80      90      100     110     120
      orf24ng      IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVVPCVPQTLKPISSRMRAATESP
      |||||
      orf24ng      IIPSSSETGINAPLKPPTALEAIMPPFFFTASFSNAKAAVVPCVPQTLKPISSRMRAATESP
60      orf24ng      70      80      90      100     110     120

```

-390-

		130	140	150	160	170	180
	orf24-1.pep	TAGVGASDKSRIPNGIFSI FEASRPMS SPTRVILKAVFFTT SATSVNVVASEFSNA AFTT					
	orf24ng	TAGVGASDKSRMPNGIFSI FEASRPMS SPTRVILKAVFFTT SATSVRLTASEFSSAALT					
5		130	140	150	160	170	180
	orf24-1.pep	PGPDTPTLTITASASPEPXNAPAINGLSSTALQNTTILAQPKPSGVISAVRLTVSPASLTA					
	orf24ng	PGPDTPTLTITASASPEPWNAPAINGLSSTALQNTTILAQPKPSGVISAVRLMVSPASLTA					
10		190	200	210	220	230	240
	orf24-1.pep	SILIPARVLPILMELHTISVVFIA SGMERINTSSEGDIPFCTNAEKPPIKDTPMALAALS					
	orf24ng	SILIPARVLPILMELHTISVVFIA SGTERINTSSEGDIPFCTSAEKPPIKDTPMALAALS					
15		250	260	270	280	290	300
	orf24-1.pep	KVCATLTX					
	orf24ng	KVCATLTX					
20		250	260	270	280	290	300

Based on this analysis, including the presence of a putative leader sequence (first 18 aa – double-
underlined) and putative transmembrane domains (single-underlined) in the gonococcal protein,
it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could
be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 81

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 681>:

30	1	..ACCGACGTGC	AAAAAGAGTT	GGTCGCGGAA	CAACGCAAGT	GGGCGCAGGA
	51	AAAAATCAGC	AACTGCCGAC	AAGCCGCCGC	GCAGGCAGAC	CGGCAGGAAT
	101	ACGCCGAATA	CCTCAAGCTG	CAATGCGACA	CGCGGATGAC	GCGCGAACGG
	151	ATACAGTATC	TTCGCGGCTA	TTCCATCGAT	TAG	

This corresponds to the amino acid sequence <SEQ ID 682; ORF25>:

35	1	..TDVQKELVGE	QRKWAQEKIS	NCRQAAAQAD	RQEYAEYLKL	QCDTRMTRER
	51	IQYLRGYSID	*			

Further work revealed the complete nucleotide sequence <SEQ ID 683>:

	1	ATGTATCGGA	AACCTATTGC	GCTGCCGTTT	GCCCTGCTGC	TTGCCGCTTG
	51	CGGCAGGGAA	GAACCGCCCA	AGGCATTGGA	ATGCGCCAAC	CCCGCCGTGT
40	101	TGCAAGGCAT	ACGCGGCAAT	ATTCAAGAAA	CGCTCAGCA	GGAAGCGCGT
	151	TCTTTCGCGC	GCGAAGACGG	CAGGCAGTTT	GTCGATGCCG	ACAAAATTAT
	201	CGCCGCCGCC	TACGGTTTGG	CGTTTCTTT	GGAACACGCT	TCGGAACACG
	251	AGGAAGGCGG	GCGCACGTTT	TGTATCGCCG	ATTTGAACAT	TACCGTGCCG
	301	TCTGAAACGC	TTGCCGATGC	CAAGGCAAAAC	AGCCCCCTGT	TGTACGGGGA
45	351	AACTGCTTTG	TCGGATATTG	TGCGGCAGAA	GACGGGCGGC	AATGTCGAGT
	401	TTAAAGACGG	CGTATTGACG	GCAGCCGTCC	GCTTCTTGCC	CGTCAAGAC
	451	GGTCAGACGG	CATTTGTCGA	CAACACGGTC	GGTATGGCGG	CGCAAACGCT
	501	GTCTGCCCGC	CTGCTGCCTT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
	551	GCAAGGCGGT	GAAAAAGAA	GACGCGGTCA	GGATTTTGAG	CGGAAAAGCC
50	601	CGTGAAGAAG	AACCGTCCAA	ACCCACGCC	GAAGACATT	TGGAACACAA
	651	TGCCGCCGCG	GGCGATGCGG	GCGTACCCCA	AGCCGCAGAA	GGCGCGCCCG
	701	AACCGGAAAT	CCTGCATCCT	GACGACGGCG	AGCGTGCCGA	TACCGTTACC
	751	GTATCACGGG	GCGAAGTGGA	AGAGGCGCGC	GTACAAAACC	AGCGTGCCGA
55	801	ATCCGAAATT	ACCAAACTTT	GGGGAGGACT	CGATACCGAC	GTGCAAAAAG
	851	AGTTGGTCGG	CGAACAACGC	AAGTGGGCGC	AGGAAAAAAT	CAGCAACTGC
	901	CGACAAGCCG	CCGCGCAGGC	AGACCGGCAG	GAATACGCCG	AATACCTCAA
	951	GCTGCAATGC	GACACGCGGA	TGACGCGCGA	ACGGATACAG	TATCTTCGCG
	1001	GCTATTCCAT	CGATTAG			

This corresponds to the amino acid sequence <SEQ ID 684; ORF25-1>:

```

1  MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQIGIRN IQETLTQEAR
51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP
101 SETLADAKAN SPILLYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
151 GQTAFVDNTV GMAAQTLCAA LPPYGVKSIV MIDGKAVKKE DAVRILSGKA
201 REEPPSKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDGERADTVT
251 VSRGEVEEAR VQNRASEI TKLWGLDLD VQKELVGEQR KWAQEKISNC
301 RQAAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID*

```

Computer analysis of this amino acid sequence gave the following results:

10 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF25 shows 98.3% identity over a 60aa overlap with an ORF (ORF25a) from strain A of *N.*

meningitidis:

```

15 orf25.pep                                10      20      30
                                         TDVQKELVGEQRKWAQEKISNCRQAAAQAD
                                         |||
orf25a      VTVSRGEVEEARVQNRASEITKLWGLDLDVQKELVGEQRKWAQEKISNCRQAAAQAD
              250      260      270      280      290      300

20 orf25.pep              40      50      60
                                         ROEYAEYLKLCQDTRMTRERIQYLRGYSIDX
                                         |||
orf25a      ROEYAEYLKLCQDTRMTRERIQYLRGYSIDX
              310      320      330

```

The complete length ORF25a nucleotide sequence <SEQ ID 685> is:

```

25 1  ATGTATCGGA AACTCATTGC GCTGCCGTTT GCCCTGCTGC TTGCCGCTTG
51 CGGCAGGGAA GAACCGCCCA AGGCATTGGA ATGCGCCAAC CCCGCCGTGT
101 TGCAANGCAT ACGCNGCAAT ATTCAGGAAA CGCTCAGCA GGAAGCGCGT
151 TCTTTCGCGC GCGAAGACNG CANGCAGTTT GTCGATGCCG ACNAAATTAT
201 CGCCGCCGCC TANGNTNNGN NGNTNTCTTT GGAACACGCT TCGGAAACGC
301 AGGAAGGCGG GCGCACGTTT TGTNTCGCCG ATTTGAACAT TACCGTGCCG
351 TCTGAAACGC TTGCCGATGC CAAGGCAAAAC AGCCCCCTGC TGTACGGGGA
401 AACCGCTTTG TCGGATATTG TCGGCGAGAA GACGGGCGGC AATGTCGAGT
451 TTAAGACGGG CGTATTGACG GCAGCGTCC GCTTCTACC CGTCAAAGAC
501 GGTGAGANGG CATTTGTCGA CAACACGGTC GGTATGGCGG CGCAAACGCT
35 GTCTGCCGCG TTGCTGCCTT ACGGCGTGAA GAGCATCGTG ATGATAGACG
551 GCAAGGCGGT AAAAAAGAA GACGCGGTCA GGATTNTGAG CNGANAAGCC
601 CGTGAANAAG AACCGTCCA ANCCNNGCCC GAAGACATT TGGAAACATAA
651 TGCCGCCGGA GGGGATGCAG ACGTACCCA AGCCGGAGAA GACGCGCCCG
701 AACCGGAAAT CCTGCATCCT GACGACGCG AGCGTGCCGA TACCGTTACC
40 751 GTATCACGGG GCGAAGTGA AGAGGCGCGN GTACAAAACC AGCGTGCCGA
801 ATCCGAAATT ACCAACTTT GGGGAGGACT CGATACCGAC GTGCAAAAAG
851 AGTTGGTCGG CGAANAACGC AAGTGGGCGC AGGAAAAAT CAGCAACTGC
901 CGACAAGCCG CCGCGCAGGC AGACCGGCAG GAATACGCCG AATACCTCAA
951 GCTGCAATGC GACACGCGGA TGACGCGCGA ACGGATACAG TATCTTCGCG
45 1001 GCTATTCCAT CGATTAG

```

This encodes a protein having amino acid sequence <SEQ ID 686>:

```

1  MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQXIRXN IQETLTQEAR
51 SFAREDXXQF VDADKIIAAA XXXXSLEHA SETQEGGRTF CXADLNITVP
101 SETLADAKAN SPILLYGETAL SDIVRQKTGG NVEFKDGVLT AAVRFLPVKD
151 GQXAFVDNTV GMAAQTLCAA LPPYGVKSIV MIDGKAVKKE DAVRIXSXXA
201 REXEPSKXXP EDILEHNAAG GDADVPQAGE DAPEPEILHP DDGERADTVT
251 VSRGEVEEAR VQNRASEI TKLWGLDLD VQKELVGEQR KWAQEKISNC
301 RQAAAQADRQ EYAEYLKLCQ DTRMTRERIQ YLRGYSID*

```

ORF25a and ORF25-1 show 93.5% identity in 338 aa overlap:

```

55 orf25a.pep              10      20      30      40      50      60
                                         MYRKLIALPFALLLAACGREPPKALECANPAVLQXIRXNIQETLTQEARSFAREDXXQF
                                         |||
orf25-1      MYRKLIALPFALLLAACGREPPKALECANPAVLQIGIRNIQETLTQEARSFAREDGRQF

```

-392-

		10	20	30	40	50	60
		70	80	90	100	110	120
5	orf25a.pep	VDADKIIAAAXXXXXXSLHASETQEGGRTFCXADLNITVPSETLADAKANSPLLYGETAL					
	orf25-1	VDADKIIAAAYGLAFSLHASETQEGGRTFCIADLNITVPSETLADAKANSPLLYGETAL					
		70	80	90	100	110	120
10	orf25a.pep	130	140	150	160	170	180
	orf25-1	SDIVRQKTGGNVEFKDGVLTAAVRFLPVKDGQXAFVDNTVGMMAAQTLSAALLPYGVKSIV					
		130	140	150	160	170	180
15	orf25a.pep	190	200	210	220	230	240
	orf25-1	MIDGKAVKKEDAVRIXSXXAREXEPSKXXPEDILEHNAAGGDADVPOAGEDAPEPEILHP					
		190	200	210	220	230	240
20	orf25a.pep	250	260	270	280	290	300
	orf25-1	DDGERADTVTVSRGEVEEARVQNQRAESEITKLWGGLDQVQKELVGEQRKWAQEKISNC					
25		250	260	270	280	290	300
	orf25a.pep	310	320	330	339		
	orf25-1	RQAAAQADRQEYAEYLKLCQDTRMTREIRIQYLRGYSID					
30		310	320	330			

Homology with a predicted ORF from *N.gonorrhoeae*

ORF25 shows 100% identity over a 60aa overlap with a predicted ORF (ORF25ng) from

35 *N.gonorrhoeae*:

	orf25.pep	TDVQKELVGEQRKWAQEKISNCRQAAAQAD	30
	orf25ng	VTVSRGEVEEARVQNQRAESEITKLWGGLDQVQKELVGEQRKWAQEKISNCRQAAAQAD	308
40	orf25.pep	RQEYAEYLKLCQDTRMTREIRIQYLRGYSID	60
	orf25ng	RQEYAEYLKLCQDTRMTREIRIQYLRGYSID	338

The complete length ORF25ng nucleotide sequence <SEQ ID 687> is:

45	1	ATGTATCGGA	AACATTCG	GCTGCCGTTT	GCCCTGCTGC	TTGCAGCGTG
	51	CGGCAGGGAA	GAACCGCCCA	AGGCGTTGGA	ATGCGCCAAC	CCCGCCGTGT
	101	TGCAGGACAT	ACGCGGCAGT	ATTCAGGAAA	CGCTCAGCA	GGAAGCCGCT
	151	TCTTTCGCGC	GCGAAGACGG	CAGGCAGTTT	GTCGATGCCG	ACAAAATTAT
	201	CGCCGCGGCC	TACGGTTTGG	CGTTTTCTTT	GGAACACGCT	TCGGAACCGC
	251	AGGAAGGCGG	GCGCAGCTTC	TGTATCGCCG	ATTGAAACAT	TACCGTGCCG
50	301	TCTGAAACGC	TTGCCGATGC	CGAGGCAAAC	AGCCCCCTGC	TGTATGGGGA
	351	AACGTCTTTG	GCAGACATCG	TGCAGCAGAA	GACGGGCGGC	AATGTCGAGT
	401	TTAAAGACGG	CGTATTGACG	GCAGCCGTCC	GCTTCCTGCC	CGCCAAAGAC
	451	GCTCGGACGG	CATTATTCGA	CAACACGGTC	GGTATGGCGA	CGCAAACGCT
	501	GTCTGCCGCG	TTGCTGCCTT	ACGGCGTGAA	GAGCATCGTG	ATGATAGACG
55	551	GCAAGGCGGT	GACAAAAGAA	GACGCGGTCA	GGGTTTTGAG	CGGCAAAGCC
	601	CGTGAAGAAG	AACCGTCCAA	ACCCACCCCC	GAAGACATTT	TGGAACACAA
	651	TGCCGCCGGC	GGCGATGCGG	GCGTACCCCA	AGCCGCAGAA	GGCGCACCCG
	701	AACCCGAAAT	CCTGCATCCC	GACGACGTCG	AGCGTGCCGA	TACCGTTACC
	751	GTATCACGGG	GCGAAGTGGA	AGAGGCGCGC	GTACAAAACC	AACGTGCGGA
60	801	ATCCGAAATT	ACCAAATTTT	GGGGAGGACT	CGATACCGAC	GTGCAAAAAG
	851	AGTTGGTCGG	CGAACAGCGC	AAGTGGGCGC	AGGAAAAAAT	CAGcaactgc
	901	cgACAAGCCG	CCGCGCAGGC	AGACCGGCAG	GAATACGCCG	AATACCTCAA
	951	GCTCCAATGC	GACACGCGGA	TGACGCGCGA	ACggaTACAG	TATCTTCGCG
	1001	GCTATTCCAT	CGATTAG			

65 This encodes a protein having amino acid sequence <SEQ ID 688>:

-393-

1 MYRKLIALPF ALLLAACGRE EPPKALECAN PAVLQDIRGS IQETLTQEAR
 51 SFAREDGRQF VDADKIIAAA YGLAFSLEHA SETQEGGRTF CIADLNITVP
 101 SETLADAEAN SPLLYGETSL ADIVQKTTGG NVEFKDGVLT AAVRFLPAKD
 151 ARTAFIDNTV GMATQTLCAA LLYGVKSIV MIDGKAVTKE DAVRVLGSKA
 201 REEPPSKPTP EDILEHNAAG GDAGVPQAAE GAPEPEILHP DDVERADTVT
 251 VSRGEVEEAR VQNQRAESEI TKLWGGLDTD VQKELVGEQR KWAQEKISNC
 301 RQAAAQADRQ EYAEYLKLCQ DTRMTREIRIQ YLRGYSID*

ORF25ng and ORF25-1 show 95.9% identity in 338 aa overlap:

10	orf25-1.pep	10	20	30	40	50	60
	orf25ng	10	20	30	40	50	60
15	orf25-1.pep	70	80	90	100	110	120
	orf25ng	70	80	90	100	110	120
20	orf25-1.pep	130	140	150	160	170	180
	orf25ng	130	140	150	160	170	180
25	orf25-1.pep	190	200	210	220	230	240
	orf25ng	190	200	210	220	230	240
30	orf25-1.pep	250	260	270	280	290	300
	orf25ng	250	260	270	280	290	300
35	orf25-1.pep	310	320	330	339		
	orf25ng	310	320	330	339		

45 Based on this analysis, including the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site (underlined) in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF25-1 (37kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 16A shows the results of affinity purification of the GST-fusion protein, and Figure 16B shows the results of expression of the His-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 16C), ELISA (positive result), and FACS analysis (Figure 16D). These experiments confirm that ORF25-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 16E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF25-1.

Example 82

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 689>

```

      1 ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
5      51 TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
      101 GCATCGGTAT TCTGGWysGC GTTGCCCTTTT TGGTCGGCGG CAACCCCGTC
      151 GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
      201 CGsyGATTGG TCGCTGGGCA AACCAAAAT CTTGGTTTTC CkGATACTTT
      251 TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA T.....
                                     //
      851 .....AC TTCGCTGGTA
      901 TTCGGCGGCA CTTGCGGCGT CTTTGCCGTC GTTCTCTGCA CGCTCGGCAC
      951 GATTAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
1001 TGTTCCGGCGC AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
1051 GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
1101 CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGCTC GCCAGCGTGA
1151 TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
1201 ATTGCCGCGG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
1251 TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
1301 TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC
1351 GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCGC
1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
1451 TTGGCACGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTGAAAGAT
1501 AAAAAA..

```

25 This corresponds to the amino acid sequence <SEQ ID 690; ORF26>:

```

      1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILXX VAFVGGNPV
51      DGLTHLKMV VGLAWSDXDW SLGKPKILVF XILLGIFTSL LTYSGSN...
                                     //
251 .....TSLV
301 FGGTCGVFAV VLCTLGITKT ADYPKAVWQG AKSMFGAIAI LILAWLISTV
351 VGEMHTGDYL STLAVAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARNHI
451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD
501 KK..

```

35 Further work revealed the complete nucleotide sequence <SEQ ID 691>:

```

      1 ATGCAGCTGA TCGACTATTC ACATTCATTT TTCTCGGTTG TGCCACCCTT
51      TTTGGCACTG GCACTTGCCG TCATTACCCG CCGCGTACTG CTGTCTTTAG
101      GCATCGGTAT TCTGGTCGGC GTTGCCCTTTT TGGTCGGCGG CAACCCCGTC
151      GACGGTCTGA CACACCTGAA AGACATGGTC GTCGGCTTGG CTTGGTCAGA
201      CGGCGATTGG TCGCTGGGCA AACCAAAAT CTTGGTTTTC CTGATACTTT
251      TGGGTATTTT TACTTCCCTG CTGACCTACT CCGGCAGCAA TCAGGCGTTT
301      GCCGACTGGG CAAAACGGCA CATTAAAAAC CGGCGCGGCG CGAAAAATGCT
351      GACCGCCTGC CTCGTGTTCG TAACCTTTAT CGACGACTAT TTCCACAGTC
401      TCGCCGTCGG TCGGATTGCC CGCCCCGTTA CCGACAAGTT TAAAGTTTCC
451      CGCACCAAAC TCGCCTACAT CCTCGACTCC ACTGCCGCTC CTATGTGCGT
501      GCTGATGCCC GTTCAAGCT GGGGCGCGTC GATTATCGCC ACGCTTGCCG
551      GACTGCTCGT TACCTACAAA ATCACCGAAT ACACGCCGAT GGGGACGTTT
601      GTCGCCATGA GCCTGATGAA CTATTACGCA CTGTTTGCCC TGATTATGGT
651      GTTCGTCGTC GCATGGTTT CTTCGACAT CGGCTCGATG GCACGTTTCG
701      AACAAAGCCG GTTGAACGAA GCCACGATG AACTGCGGT TTCAGACGCT
751      ACCAAAGGTC GTGTTTACGC ACTGATTATT CCCGTTTGG CCTTAATCGC
801      CTAACCGGTT TCCGCCATGA TCTACACCGG CGGCGAGGCA AGCGAAACCT
851      TCAGCATTTT GGGGCGATT GAAAACACGG ACGTAAACAC TTCGCTGGTA
901      TTCGGCGGCA CTTGCGGCGT CCTTGCCGTC GTTCTCTGCA CGCTCGGCAC
951      GATTAAACC GCCGACTATC CCAAAGCCGT TTGGCAGGGT GCGAAATCTA
1001      TGTTCCGGCG AATCGCCATT TTAATCCTCG CTTGGCTCAT CAGTACGGTT
1051      GTCGGCGAAA TGCACACCGG CGATTACCTC TCCACACTGG TTGCGGGCAA
1101      CATCCATCCC GGCTTCCTGC CCGTCATCCT CTTCTGCTC GCCAGCGTGA
1151      TGGCGTTTGC CACAGGCACA AGCTGGGGGA CGTTCGGCAT TATGCTGCCG
1201      ATTGCCGCG CCATGGCGGT CAAAGTCGAA CCCGCGCTGA TTATCCCGTG
1251      TATGTCCGCA GTAATGGCGG GGGCGGTATG CGGCGACCAC TGCTCGCCCA
1301      TTTCCGACAC GACCATCCTG TCGTCCACCG GCGCGCGCTG CAACCACATC

```

```

1351 GACCACGTTA CCTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC
1401 CGCATCGGGC TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGCT
1451 TTGGCACGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAAGAT
1501 AAAAAACGCG CCAACGCCTG A

```

5 This corresponds to the amino acid sequence <SEQ ID 692; ORF26-1>:

```

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVGGNPV
51 DGLTHLKDMV VGLAWSDDGW SLGKPKILVF LILLGIFTSL LTYSGSNQAF
101 ADWAKRHIKN RRGAKMLTAC LVFVTFIDY FHS LAVGAIA RPVTDKFKVS
151 RTKLAYILDS TAAPMCVLMF VSSWGASIIA TLAGLLVYK ITEYTPMGTF
10 VAMSLMNYA LFALIMVFVV AWFSFDIGSM ARFEQAALNE AHD EAVSDA
251 TKGRVYALII PV LALIASTV SAMIYTGAA SETFSILGAF ENT D VNTSLV
301 FGGTCGVLA V LCTLTG IKT ADY PKAVWQ AKSMFGAIAI LILAWLISTV
351 VGEMHTG DY L STL VAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARNHI
15 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD
501 KKRANA*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical transmembrane protein HI1586 of *H.influenzae* (accession number P44263)

ORF26 and HI1586 show 53% and 49% amino acid identity in 97 and 221 aa overlap at the

20 N-terminus and C-terminus, respectively:

```

Orf26 1 MQLIDYSHSFFSVVPPFLALALAVITRRVXXXXXXXXXXVAFVGGNPVDGLTHLKDMV 60
M+LID+S S +S+VP LA+ LA+ TRRV L +L V
HI1586 14 MELIDFSSSVSIVPALLAIILAIATRRVLVSLSAGIIGSLMLSDWQIGSAFNVLVKNV 73

25 Orf26 61 VGLAWSDXDWSLGKPKILVFXILLGIFTSLTYSGSN 97
V L ++D + + I++F +LLG+ T+LLT SGSN
HI1586 74 VSLVYADGEIN-SNMNIVLFLLLLGVLTALLTVSGSN 109

//

30 Orf26 86 IFTSLLTYSGS--NTSLVFGGTCGVFAVVLCTL--GTIKTADY PKAVWQAKSMFGXXXX 141
+F+ L T+ + TSLV GG C + L + + +Y ++ G KSM G
HI1586 299 VFSVLGTFENTVVGTSLVVGGFCSSIIISTLLIILDRQVSVPYVRSWIVG+KSMGSAIAI 358

35 Orf26 142 XXXXXXSTVVGEMHTGDY LSTLVAGNIHPGFLPVILFLLASVMAFATGT SWGTFGIMLP 201
+ +VG+M TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLP
HI1586 359 LFFAWTINKIVGDMQTGKYLSSLVSGNIPMQFLPVILFVLGAAMAFSTGT SWGTFGIMLP 418

40 Orf26 202 IAAAMAVKVEPALIIPCMSAVMAGAVCGDH CSPISDTTILSSTGARNHIDHVTSQXXXX 261
IAAAMA P L++PC+SAVMAGAVCGDH CSP+SDTTILSSTGA+CNHIDHVT+Q
HI1586 419 IAAAMAANAPEL LLLPCLSAVMAGAVCGDH CSPVSDTTILSSTGAKCNHIDHVTQLPYA 478

Orf26 262 XXXXXXXXXXXXXXXXSALLGFGTTGIVLAVLIFLLKDK 302
S L GF T + L V+IF +K +
45 HI1586 479 ATVATATSIGYIVVGFTYSGLAGFAATAVSLIVIFAVKKR 519

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF26 shows 58.2% identity over a 502aa overlap with an ORF (ORF26a) from strain A of *N.*

meningitidis:

```

50 orf26.pep 10 20 30 40 50 60
MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFVGGNPVDGLTHLKDMV
|||||
orf26a 10 20 30 40 50 60
MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV

55 orf26.pep 70 80 90 99
VGLAWSDXDWSLGKPKILVFXILLGIFTSLTYSGSNXX-----
|||||
orf26a 70 80 90 100 110 120
VGLAWSDDWSLGKPKXLVFLILLGIFTSLTYSGSNQAFADWAKRHIKNRRGAKMLTAC

```

45	1	ATGCAGCTGA	TCGACTATT	ACATTCATTT	TTCTCGGTTG	TGCCACCCTT
	51	TTTGGCACTG	GCACTTGCCG	TCATTACCCG	CCGCGTACTG	CTGTCTTTAG
	101	GCATCGGTAT	TCTGGTCGGC	GTTGCCTTTT	TGGTCGGCGG	CAACCCCGTG
	151	GACGGTCTGA	CACACCTGAA	AGACATGGTC	GTCGGCTTGG	CTTGGTCAGA
	201	CGCGCAATTGG	TCGCTGGGCA	AACCAAAANT	CTTGGTTTTT	CTGATACTTT
	251	TGGGTATTTT	TACTTCCCTG	CTGACCTACT	CCGGCAGCAA	TCAGGCGTTT
50	301	GCCGACTGGG	CAAAACGGCA	CATTAAAAAC	CGGCGCGGGC	CGAAATGCT
	351	GACCGCTGCG	CTCGTGTCG	TAACCTTTAT	CGACGACTAT	TTCCACAGTC
	401	TGCGCGTCGG	TGCGNTTGCC	CGCCCCGTTA	CCGACAAGTT	TAAAGTTTCC
	451	CGCGCCAAAC	TCGCTACAT	CCTCGACTCG	ACTGCCGGCG	CTATGTGCGT
	501	GCTGATGCCC	GTTTCAAGCT	GGGGCGCGTC	GATTATCGCC	ACGCTTGCCG
55	551	GACTGCTCGT	TACCTACAAA	ATCACCGAAT	ACACGCCGAT	GGGGACGTTT
	601	GTCGCCATGA	GCCTGATGAA	CTATTACGCA	CTGTTTGCCC	TGATTATGCT
	651	GTTCTGCTGTC	GCATGGTTCT	CCTTCGACAT	GGCGCTCGAT	GCACGTTTCC
	701	AACAAGCCGC	GTTGAACGAA	GCCCACGATG	AAACTGCCGT	TTCAGACGGC
	751	AGCTGGGGCA	GGGTTTACGC	ATTGATTATT	CCCGTTTTGG	CCTTAATCGC
60	801	CTCAACGGTT	TCCGCCATGA	TCTACACCGG	TGCACAGGCA	AGCGAAACCT
	851	TCAGCATTTT	GGGTGCATTT	GAAAATACGG	ACGTGAACAC	TTGCTGGTA
	901	TTCCGGCGCA	CTTGCGCGGT	GCTTGCCGTC	GTCCTCTGCA	CGCTCGGCAC
	951	GATTAAAAATC	GCCGATTATC	CCAAAGCCGT	TTGGCAGGGT	GCGAAATCCA
	1001	TGTTCCGCGC	AATCGCCATT	TTAATCCTTG	CCTGGCTCAT	CAGTACGGTT
65	1051	GTCGGCGAAA	TGCACACAGG	CGACTACCTC	TTCCAGCTGG	TTGCGGGCAA
	1101	CATCATCTCC	GGCTTCCTGN	CGCTCATCCT	TTTCTGCTC	GCCAGCGTGA
	1151	TGGCGTTTGC	CACAGGCACA	AGCTGGGGGA	CGTTCCGGCAT	CATGCTGCCG
	1201	ATTGCCGCCG	CCATGGCGGT	CAAAATCGAT	CCCTCACTGA	TTATCCCGTG
	1251	TATGTCGGCG	GTGATGGCGG	GGCGGGTATG	CGCGGACCAC	TGCTCGCCCA
70	1301	TTTCCGACAC	GACCATCTCG	TCGTCCACCG	GCGCGCGCTG	CAACCACATC

1351 GACCACGTTA CNTCGCAACT GCCTTACGCC TTAACCGTTG CCGCCGCCGC
 1401 CGCATCGGGN TACCTCGCAT TGGGTCTGAC AAAATCCGCG CTGTTGGGTT
 1451 TTGGCANGAC AGGCATTGTA TTGGCGGTGC TGATTTTCT GTTGAAGAT
 1501 AAAAAACGCG CCAACGCTG A

5 This encodes a protein having amino acid sequence <SEQ ID 694>:

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFVGGNPV
 51 DGLTHLKMV VGLAWSGDW SLGKPKXLVF LILGIFTS LTYSGSNQAF
 101 ADWAKRHIK RRGAKMLTAC LVFVTFIDY FHSLAVGAXA RPVTDKFKVS
 151 RAKLAYILDS TAAPMCVLMF VSSWGASIIA TLAGLLVYK ITEYTPMGTF
 10 VAMSLMNYA LFALIMVFVV AWFSFDIGSM ARFEQAALNE AHDEAVSDG
 251 SWGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTDVNTSLV
 301 FGGTCGVLAV VLCTLGTIKI ADYPKAVWQ AKSMFGAIAI LILAWLISTV
 351 VGEMHTGDYL STLVAGNIHP GFLXVILFLL ASVMAFATGT SWGTFGIMLP
 401 IAAAMAVKVD PSLIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARCNIH
 15 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGXTGIV LAVLIFLLKD
 501 KKRANA*

ORF26a and ORF26-1 show 97.8% identity in 506 aa overlap:

20	orf26a.pep	10	20	30	40	50	60
	orf26-1	10	20	30	40	50	60
25	orf26a.pep	70	80	90	100	110	120
	orf26-1	70	80	90	100	110	120
30	orf26a.pep	130	140	150	160	170	180
	orf26-1	130	140	150	160	170	180
35	orf26a.pep	190	200	210	220	230	240
	orf26-1	190	200	210	220	230	240
40	orf26a.pep	250	260	270	280	290	300
	orf26-1	250	260	270	280	290	300
50	orf26a.pep	310	320	330	340	350	360
	orf26-1	310	320	330	340	350	360
55	orf26a.pep	370	380	390	400	410	420
	orf26-1	370	380	390	400	410	420
60	orf26a.pep	430	440	450	460	470	480
	orf26-1	430	440	450	460	470	480
65	orf26a.pep	490	500				
	orf26a.pep	LLGFGXTGIVLAVLIFLLKDKKRANAX					

5 Homology with a predicted ORF from *N.gonorrhoeae*

ORF26 shows 94.8% and 99% identity in 97 and 206 aa overlap at the N-terminus and C-terminus, respectively, with a predicted ORF (ORF26ng) from *N. gonorrhoeae*:

	orf26.pep	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILXXVAFVLVGGNPVDGLTHLKDMV	60
10	orf26ng	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFVLVGGNPVDGLTHLKDMV	60
	orf26.pep	VGLAWSDXDWSLGGPKILVFXILLGIFTSLLTYSGSN	97
15	orf26ng	VGLAWADGDWSLGGPKILVFLILLGIFTSLLTYSGSNQAFADWAKRHIKNRCGAKMLTAC	120
		//	
	orf26.pep	TSLVFGGTGCGVFAVVLCTLGTIKTADYPKA	326
20	orf26ng	ASTVSAMIYTGAQASETFSILGAFENTDVNTSLVFGGTGCGVLAVVLCTFGTIKTADYPKA	326
	orf26.pep	VWQGAKSMFGAIAILILAWLISTVVGEMHTGDYLSLTVAGNIHPGFLPVILFLLASVMAF	386
25	orf26ng	VWQGAKSMFGAIAILILAWLISTVVGEMHTGDYLSLTVAGNIHPGFLPVILFLLASVMAF	386
	orf26.pep	ATGTSWGTFGIMLPPIAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISDTTILSSTGAR	446
30	orf26ng	ATGTSWGTFGIMLPPIAAAMAVKVEPALIIPCMSAVMAGAVCGDHCSPISDTTILSSTGAR	446
	orf26.pep	CNHIDHVTSQLPYALTVA AAAASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKK	502
	orf26ng	CNHIDHVTSQLPYALTVA AAAASGYLALGLTKSALLGFGTTGIVLAVLIFLLKDKK RADV	506

The complete length ORF26ng nucleotide sequence <SEQ ID 695> is:

	1	ATGCAGCTGA	TTGACTATTG	ACATTCATT	TTCTCGGTTG	TGCCACCCTT
35	51	TTTGCGACTG	GCATTGCGG	TCATTACCGG	CCGCGTACTG	CTGCTCTTAG
	101	GCATCGGTAT	TTTGGTCCGG	GTTGCCTTTT	TGGTGCGGCG	CAACCCCGTC
	151	GACGGTCTGA	CACACCTGAA	AGACATGGTC	GTGCGGCTGG	CTTGGGCAGA
	201	CGGCGATTGG	TCGCTGGGCA	AACCAAAAAT	CTTGGTTTTT	CTGATACTTT
40	251	TGGGCATTTT	CACCTTCACTG	CTGACCTACT	CCGGCAGCAA	TCAGGCGTTT
	301	GCCGACTGGG	CAAAACGGCA	CATTA AAAAC	CGGTGCGGCG	CGAAAATGCT
	351	GACCCGCTGC	CTCGTGTTTCG	TAACTTTAT	TCACGACTAT	TTCCACAGCC
	401	TCGCGCTCGG	TGCGATTGCC	CGCCCCGTTA	CCGACAAGTT	TAAAGTTTCC
	451	CGCGCCAAAC	TCGCCTACAT	CCTCGACTCC	ACTGCCTCGC	CCATGTGCGT
45	501	GCTGATGCCC	GTTTCAAGCT	GGGGCGCGTC	GATTATCGCC	ACGCTTGCCG
	551	GATTGCTCGT	TACCTACAAA	ATTACCGAAT	ACACGCGCAT	GGGGACGTTT
	601	GTCGCCATGA	GCCTGATGAA	CTATTACCGG	CTGTTTGCCC	TGATTATGGT
	651	ATTCTGTCGT	GCATGGTTCT	CCTTCGACAT	CGGCTCGATg	gCGCGTTTCG
	701	AACAGGCTGC	GTGTAACGAA	gccccaggagc	aaaccgcgcg	tTCAGACgCT
	751	ACCAAAGGTC	GTTTATACGC	ATTGATTATT	CCCGTTTTTG	CCTTAATCGC
50	801	CTCAACGGTT	TCCGCCATGA	TCTACACCGG	CGCGCAGGCA	AGCGAAACCT
	851	TCAGCATTTT	GGGGGCATT	GAAAATACCG	ACGTAACAC	TTGCTGGTA
	901	TTGCGCGGCA	GTTGCGCGT	GCTTGCCGTC	GTCTCTGCA	CGTTCGGCAC
	951	GATTA AAACC	CCTGATTATC	CCAAAGCCGT	GTGGCAGGGT	GCGAAATCCA
	1001	TGTTCGGCGC	AATCGCCATT	TTAATCCTCG	CCTGGCTCAT	CAGTACGGTT
55	1051	GTGCGCGAAA	TGCACACGGG	CGACTACCTC	TCCACGCTGG	TTGCGGGCAA
	1101	CATCCATCCC	GGCTTCTTGC	CCGTATCCT	CTTCTGTCT	GCCAGCGTGA
	1151	TGGCGTTTGC	CACAGGCACA	AGCTGGGGGA	CGTTCCGGCAT	TATGCTGCCG
	1201	ATTGCCGCGG	CCATGGCGGT	CAAAGTCGAA	CCCGCGCTGA	TTAtcccGTG
	1251	TATGTCCGCA	GTAATTGGCGG	GGGCGGTATG	CGGCGACCAC	TGTTCCGCCA
60	1301	TCTCCGACAC	GACCATCTTG	TCGTCACACG	GCGCGCGCTG	CAACCACATC
	1351	GACCACGTTA	CCTCGCAACT	GCCTTATGCC	CTGACGGTTG	CCGCGCCGCG
	1401	CGCATCGGGC	TACCTCGCAT	TGGGTCTGAC	AAAATCCGCG	CTGTTGGGCT
	1451	TTGGCACGAC	CGGTATTGTA	TTGGCGGTGC	TGATTTTTCT	GTTGAAAGAT
	1501	AAAAAACGCG	CCGACGTTTG	A		

65 This encodes a protein having amino acid sequence <SEQ ID 696>:

-399-

1 MQLIDYSHSF FSVVPPFLAL ALAVITRRVL LSLGIGILVG VAFLVGGNPV
 51 DGLTHLKDMV VGLAWADGDW SLGKPKILVF LILLGIFTSL LTYSGSNQAF
 101 ADWAKRHIKN RCGAKMLTAC LVFVTFIDY FHSLAVGAIARPVTDFKVS
 151 RAKLAYILDS TASPVCVIMP VSSWGASIIA TLAGLLVITYK ITEYTPMGTF
 201 VAMSLMNYIA LFALIMVFFV VWFSDIGSM ARFEQAALNE AQDETAASDA
 251 TKGRVYALII PVLALIASTV SAMIYTGAQA SETFSILGAF ENTDVNTSLV
 301 FGGTCGVLAV VLCTFGTIKT ADYPAVWQG AKSMFGAIAI LILAWLISTV
 351 VGMHTGDYL STLVAAGNIHP GFLPVILFLL ASVMAFATGT SWGTFGIMLP
 401 IAAAMAVKVE PALIIPCMSA VMAGAVCGDH CSPISDTTIL SSTGARCNIH
 451 DHVTSQLPYA LTVAAAAASG YLALGLTKSA LLGFGTTGIV LAVLIFLLKD
 501 KKRAADV*

ORF26ng and ORF26-1 show 98.4% identity in 505 aa overlap:

15	orf26-1.pep	10 20 30 40 50 60	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
	orf26ng	10 20 30 40 50 60	MQLIDYSHSFFSVVPPFLALALAVITRRVLLSLGIGILVGVAFLVGGNPVDGLTHLKDMV
20	orf26-1.pep	70 80 90 100 110 120	VGLAWSDDGWSLGKPKILVFLILLGIFTSLTYSGSNQAFADWAKRHIKNRRGAKMLTAC
	orf26ng	70 80 90 100 110 120	VGLAWADGDWSLGKPKILVFLILLGIFTSLTYSGSNQAFADWAKRHIKNRRGAKMLTAC
25	orf26-1.pep	130 140 150 160 170 180	LVFVTFIDYFHSLAVGAIARPVTDFKVSRTKLAYILDSTAAPMCVIMPVSSWGASIIA
	orf26ng	130 140 150 160 170 180	LVFVTFIDYFHSLAVGAIARPVTDFKVSRAKLAYILDSTASPMCVIMPVSSWGASIIA
30	orf26-1.pep	190 200 210 220 230 240	TLAGLLVITYKITEYTPMGTFVAMSLMNYIALFALIMVFFVAVWFSDIGSMARFEQAALNE
	orf26ng	190 200 210 220 230 240	TLAGLLVITYKITEYTPMGTFVAMSLMNYIALFALIMVFFVAVWFSDIGSMARFEQAALNE
35	orf26-1.pep	250 260 270 280 290 300	AHDETAASDATKGRVYALII PVLALIASTVSAMIYTGAQASETFSILGAFENTDVNTSLV
40	orf26ng	250 260 270 280 290 300	AQDETAASDATKGRVYALII PVLALIASTVSAMIYTGAQASETFSILGAFENTDVNTSLV
45	orf26-1.pep	310 320 330 340 350 360	FGGTCGVLAVVLCTLTGTIKTADYPAVWQGAQSMFGAIAI LILAWLISTV VGMHTGDYL
	orf26ng	310 320 330 340 350 360	FGGTCGVLAVVLCTLTGTIKTADYPAVWQGAQSMFGAIAI LILAWLISTV VGMHTGDYL
50	orf26-1.pep	370 380 390 400 410 420	STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSA
	orf26ng	370 380 390 400 410 420	STLVAGNIHPGFLPVILFLLASVMAFATGTSWGTFGIMLP IAAAMAVKVEPALIIPCMSA
55	orf26-1.pep	430 440 450 460 470 480	VMAGAVCGDH CSPISDTTILSSTGARCNIH DHVTSQLPYALTVA AAAASGYLALGLTKSA
	orf26ng	430 440 450 460 470 480	VMAGAVCGDH CSPISDTTILSSTGARCNIH DHVTSQLPYALTVA AAAASGYLALGLTKSA
60	orf26-1.pep	490 500	LLGFGTTGIVLAVLIFLLKDKKRANAX
65	orf26ng	490 500	LLGFGTTGIVLAVLIFLLKDKKRADVX

In addition, ORF26 ng shows significant homology to a hypothetical *H. influenzae* protein:

-400-

sp|P44263|YF86_HAEIN HYPOTHETICAL PROTEIN HI1586 >gi|1074850|pir|C64037
 hypothetical
 protein HI1586 - Haemophilus influenzae (strain Rd KW20) >gi|1574427 (U32832) H.
 influenzae predicted coding region HI1586 [Haemophilus influenzae] Length = 519
 Score = 538 bits (1370), Expect = e-152
 Identities = 280/507 (55%), Positives = 346/507 (68%), Gaps = 7/507 (1%)

5
 10
 15
 20
 25
 30
 35
 40

Query: 1 MQLIDYSHSFFSVVPPFLALALAVITRRXXXXXXXXXXXXAFLVGGNPVDGLTHLKDMV 60
 M+LID+S S +S+VP LA+ LA+ TRR L +L V
 Sbjct: 14 MELIDFSSSVWSIVPALLAIIIAIATRRVLVSLSAGIIIGSLMLSDWQIGSAFNLYLVKNV 73

Query: 61 VGLAWADGDWSLGGPKILVFLILGIFTSLLTYSNQAFAWAKRHIKNCRAKMLTAC 120
 V L +ADG+ + I++FL+LLG+ T+LLT SGSN+AFA+WA+ IK R GAK+L A
 Sbjct: 74 VSLVYADGEIN-SNMNIVLFLLLGVLTAALLTVSGSNRAFAEWAQSRIGRRGAKLLAAS 132

Query: 121 LVFVTFIDYFHS LAVGAIARPVTDKFKVSRKALAYILDSTASPMCVLMPVSSWGASIIA 180
 LVFVTFIDYFHS LAVGAIARPVTD+FKVSRKALAYILDSTA+PMCV+MPVSSWGA II
 Sbjct: 133 LVFVTFIDYFHS LAVGAIARPVTDKFKVSRKALAYILDSTAAPMCVMMFPVSSWGAYIIT 192

Query: 181 TLAGLLVITYKITEYTPMGTFVAMSLMNYALFALIMVFVVAWFSFDIGSMARFEQAALNE 240
 + GLL TY ITEYTP+G FVAMS MN+YA+F++IMVF VA+FSFDI SM R E+ AL
 Sbjct: 193 LIGGLLATYSITEYTPIGAFVAMSSMNFYAFISIIMVFFVAYFSFDIASMVRHEKLAKN 252

Query: 241 AQDETAASDATKGRVYALIPVLALIASTVSAMIYTGAAQ-----SETFSILGAFENTDVN 296
 +D+ TKG+V LI+P+L LI +TVS MIYTGAA+ + FS+LG FENT V
 Sbjct: 253 TEDQLEETGTGQVRNLILPILVLIATVSMMIYTGAEALAADGKVSVLGTFENTVVG 312

Query: 297 TSLVFGGTCGVL--AVVLCTFGTIKTADYPKAVWQGAQSMFGXXXXXXXXXXSTVVGEM 354
 TSLV GG C ++ +++ + +Y ++ G KSM G + +VG+M
 Sbjct: 313 TSLVVGGFCSIIISTLLIILDRQVSVPEYVRSWIVGIKSMGAIAILFFAWTINKIVGDM 372

Query: 355 HTGDYLSLTVAGNIHPGFLPVILFLASVMAFATGTSWGTFGIMLPAAAMAVKVEPALI 414
 TG YLS+LV+GNI FLPVILF+L + MAF+TGTSWGTFGIMLPAAAMA P L+
 Sbjct: 373 QTGKYLSSLVSGNIIPMQFLPVILFVLAAMAFSTGTSWGTFGIMLPAAAMAANAPELL 432

Query: 415 IPCMSAVMAGAVCGDHCSPISDITLSSTGARNHIDHVT SQXXXXXXXXXXXXXXXXX 474
 +PC+SAVMAGAVCGDHCSP+SDTILSSTGA+CNHIDHVT+Q
 Sbjct: 433 LPCLSAVMAGAVCGDHCSPVSDITLSSTGAKCNHIDHVTQLPYAATVATATSIGYIVV 492

Query: 475 XXXKSALLGFGTTGIVLAVLIFLLKDK 501
 S L GF T + L V+IF +K +
 Sbjct: 493 GFTYSGLAGFAATAVSLIVIIFAVKKR 519

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*,
 45 and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 83

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 697>:

50
 1 ..AAGCAATGGT ATGCCGACGN .AGTATCAAG ACGGAAATGG TTATGGTCAA
 51 CGATGAGCCT GCCAAAATTC TGACTTGGGA TGAAAGCGGC CGATTACTCT
 101 CGGAAGTGTG TATCCGCCAC CATCAACGCA ACGGGGTGGT TTTGGAGTGG
 151 TATGAAGATG GTTCTAAAAA GAGCGAAGT. GTTTATCAGG ATGACAAGTT
 201 GGTGAGGAAA ACCCAGTGGG ATAAGGATGG TTATTTAATC GAACCTGA

This corresponds to the amino acid sequence <SEQ ID 698; ORF27>:

55
 1 ..KQWYADXSIK TEMVMVNDEP AKILTWDESG RLLSELSIRH HORNGVVLEW
 51 YEDGSKKSEX VYQDDKLVRK TQWDKDGyli EP*

Further work revealed the complete nucleotide sequence <SEQ ID 699>:

60
 1 ATGAAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC
 51 GGCCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAACGGAA
 101 AGCTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
 151 GTGGCGGGTA TTGCGCACGC GCAGGATTTT TATTATCCGT CGATGAAGAA

-401-

201 ATATTCTGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC
 251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA TGGTCAGAAA
 301 AAAATGGCGG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA
 351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGCT
 401 TGAGTGAGGG TACGGGATAC CGCTATTACC GTAACGGCGG CAAGGAAAGC
 451 GAAATCCAGT TTAAGCAAAA TAAGGCAAAC GCGGTATGGA AGCAATGGTA
 501 TGCCGACGGC AGTATCAAGA CGGAAATGGT TATGGTCAAC GATGAGCCTG
 551 CCAAAATTCT GACTTGGGAT GAAAGCGGCC GATTACTCTC GGAAGTGTCT
 601 ATCCGCCACC ATCAACGCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG
 651 TTCTAAAAAG AGCGAAGCTG TTTATCAGGA TGACAAGTTG GTCAGGAAAA
 701 CCCAGTGGGA TAAGGATGGT TATTTAATCG AACCTGA

This corresponds to the amino acid sequence <SEQ ID 700; ORF27-1>:

1 MKKLSRIVFS TVLLGFSAAL PAQYYSVYFN QNGKLTATMS SAAYIROYSV
 51 VAGIAHAQDF YPPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFNGQK
 101 KMAGGFSK GK PDGEWVNWYP NGKKSVMMPY KNGLSEGTGY RYYRNGGKES
 151 EIQFKQNKAN GVWKQWYADG SIKTEMVMVN DEPAKILTWD ESRLLSELS
 201 IRHHQRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG YLIEP*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

20 ORF27 shows 91.5% identity over a 82aa overlap with an ORF (ORF27a) from strain A of *N. meningitidis*:

						10	20	30	
	orf27.pep					KQWYADXS	SIKTEMVMVN	DEPAKILTWD	ESG
							:		
25	orf27a	LSEGTGXRYR	NGGKESEI	QFKQNKAN	GVWKQWYAD	GNIKTEMVM	VNDEPAKILT	WD	ESG
		140	150	160	170	180	190		
			40	50	60	70	80		
	orf27.pep		RLLSELSIR	HHQRNGVV	LEWYEDGSK	KSEXVYQDD	KLVRKTQWD	KDGYLIEP	X
30	orf27a		RLLSELSI	HHXRNGVV	LEWYEDGSK	KXEA	VYQDDKLVR	KTQWD	XDGYLIEP
			200	210	220	230	240		

The complete length ORF27a nucleotide sequence <SEQ ID 701> is:

1 ATGAAAAAT TATCTCGGAT TGTATTTTCA ACTGTCCTGT TGGGTTTTTC
 35 51 GGCCGCTTGT CCGGCGCAGA NCTATTCTGT TTATTTTAA CAGAACGGGA
 101 AACTGACGGC GACGNTGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
 151 GCGGAGGGTA TTGCGCACGC GCAGGANTTT TANTATCCGT CGATGAAGAA
 201 ATATTCCGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC
 251 CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA NGGTCAGAAA
 301 AAAATGGCNG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AGTGGGTCAA
 351 CTGGTATCCG AACGGTAAAA AATCTGCCGT TATGCCTTAT AAAAATGGTT
 401 TGAGTGAAGG TACGGGTTNN CGCTATTACC GTAACGGCGG CAAGGAAAGC
 451 GAAATCCAGT TTAACAGAA TAAGGCAAAC GCGGTATGGA AGCAATGGTA
 501 TGCCGACGGC AATATCAAAA CGGAAATGGT TATGGTCAAT GATGAGCCTG
 45 551 CCAAAATTCT GACATGGGAT GAAAGCGGTC GATTACTCTC GGAAGTGTCT
 601 ATCCATCATC ATNAACGTAA TGGAGTAGTC TTAGAGTGGT ATGAAGATGG
 651 TTCTAAAAAG ANTGAAGCTG TTTATCAGGA TGATAAGTTG GTCAGGAAAA
 701 CCCAGTGGGA TAANGATGGT TATTTAATCG AACCTGA

This encodes a protein having amino acid sequence <SEQ ID 702>:

50 1 MKKLSRIVFS TVLLGFSAAL PAQXYSVYFN QNGKLTATXS SAAYIROYSV
 51 AEGIAHAQXF XPPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFXGQK
 101 KMAGGFSK GK PDGEWVNWYP NGKKSVMMPY KNGLSEGTGX RYYRNGGKES
 151 EIQFKQNKAN GVWKQWYADG NIKTEMVMVN DEPAKILTWD ESRLLSELS
 201 IHXXRNGVV LEWYEDGSKK XEA

55 ORF27a and ORF27-1 show 94.7% identity in 245 aa overlap:

		10	20	30	40	50	60
	orf27a.pep	MKKLSRIVF	STVLLGFSA	ALPAQXYSV	YFNQNGKLT	ATXS	SAAYIROYS
							VAEGIAHAQ

-402-

5	orf27-1	 MKKLSRIVFSTVLLGFS AALPAQTYSVYFNQNGKLTATMSSAAYIROQSVVAGIAHAQDF 10 20 30 40 50 60
	orf27a.pep	70 80 90 100 110 120 XYPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFXGQKKMAGGFSKGPDPGEWVNWYP
10	orf27-1	70 80 90 100 110 120 YYPSMKKYSEPYIVASTQIKSFVPTLQNGMLILWHFXGQKKMAGGFSKGPDPGEWVNWYP
	orf27a.pep	130 140 150 160 170 180 NGKKS AVMPYK NGLSEGTGXRYRNGGKESEIQFKQNKANGVWKQWYADGNIKTEMVMVN
15	orf27-1	130 140 150 160 170 180 NGKKS AVMPYK NGLSEGTGYRYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVN
	orf27a.pep	190 200 210 220 230 240 DEPAKILTWDSEGRLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVKRTQWDKDG
20	orf27-1	190 200 210 220 230 240 DEPAKILTWDSEGRLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVKRTQWDKDG
	orf27a.pep	YLI EPX
25	orf27-1	YLI EPX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF27 shows 96.3% identity over 82 aa overlap with a predicted ORF (ORF27ng) from

30 *N.gonorrhoeae*:

35	orf27.pep	KQWYADXSIKTEMVMVNDEPAKILTWDSESG 30
	orf27ng	LSEGTGYRYRNGGKESEIQFKQNKANGVWKQWYADGSIKTEMVMVNDEPAKILTWDSESG 193
35	orf27.pep	RLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVKRTQWDKDG YLIEP 82
	orf27ng	RLLSELSIRHHQRNGVVLEWYEDGSKKSEAVYQDDKLVKRTQWDKDG YLIEP 245

The complete length ORF27ng nucleotide sequence <SEQ ID 703> is:

40	1	ATGAAGAAAT TATCTCGGAT TGTATTTTCA ATCGTACTGT TGGGTTTTTC
	51	GGCCGCTTTG CCGGCGCAGA CCTATTCTGT TTATTTTAAT CAGAACGGGA
45	101	AACTGACGGC GACGATGTCT TCTGCCGCTT ATATCAGGCA ATATAGTGTG
	151	GCGGCGGGTA TCGCACACGC GCAGGATTTT TATTATCCGT CGATGAAGAA
45	201	ATATTCCGAA CCTTATATCG TTGCTTCAAC GCAAATCAAA TCTTTTGTGC
	251	CTACCCTGCA AAACGGTATG TTGATTTTGT GGCATTTTAA TGGTCAGAAA
45	301	AAAATGGCGG GGGGCTTCAG CAAGGGTAAG CCGGACGGGG AATGGGTCAA
	351	CTGGTATCCG AACGGTAAAA AATCTGCGGT TATGCCTTAT AAAAATGGCT
45	401	TGAGTGAGGG TACGGGATAC CGTTATTACC GTAACGGCGG CAAGGAAAGC
	451	GAAATCCAGT TTAAGCAAAA TAAGGCGAAC GGCATATGGA AGCAATGGTA
50	501	TGCCGATGGA AGTATCAAGA CGGAAATGGT TATGGTCAAC GATGAGCCTG
	551	CCAAAATTCT GACTTGGGAT GAAAGCGGCC GATTACTTTC GGAACGTGCT
50	601	ATCCGCCACC ATAAACGCAA CGGGGTGGTT TTGGAGTGGT ATGAAGATGG
	651	TTCTAAAAAG AGCGAGGCTG TTTATCAGGA TGACAAGTTG GTCAGGAAAA
50	701	CCCAATGGGA TAAGGATGGT TATTTAATCG AACCTTGA

This encodes a protein having amino acid sequence <SEQ ID 704>:

55	1	MKKLSRIVFS IVLLGFS AAL PAQTYSVYFN QNGKLTATMS SAAYIROQSV
	51	AAGIAHAQDF YYPSMKKYSE PYIVASTQIK SFVPTLQNGM LILWHFXGQK
55	101	KMAGGFSK GK PDGEWVNWYP NGKKS AVMPY K NGLSEGTGY RYYRNGGKES
	151	EIQFKQNKAN GVWKQWYADG SIKTEMVMVN DEPAKILTWD ESGRLLSELS
55	201	IRHHKRNGVV LEWYEDGSKK SEAVYQDDKL VRKTQWDKDG YLIEP*

60 ORF27ng and ORF27-1 show 98.8% identity in 245 aa overlap:

60	orf27-1.pep	10 20 30 40 50 60 MKKLSRIVFSTVLLGFS AALPAQTYSVYFNQNGKLTATMSSAAYIROQSVVAGIAHAQDF
----	-------------	--

ORF27-1 (24.5kDa) was cloned in pET and pGex vectors and expressed in *E. coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 17A shows the results of affinity purification of the GST-fusion protein, and Figure 17B shows the results of expression of the His-fusion in *E. coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result, confirming that ORF27-1 is a surface-exposed protein and a useful immunogen.

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 705>:

50 This corresponds to the amino acid sequence <SEQ ID 706; ORF47>:

BNSDOCID: <WO 9924578A2 I >

151 HVQLHNGNLG GLLSGLQSGL VM

Further work revealed the complete nucleotide sequence <SEQ ID 707>:

```

      1 ATGAAATTTA CCAAGCACCC CGTCTGGGCA ATGGCGTTCC GCCCATTTTA
      51 TTCGCTGGCG GCTCTGTACG GCGCATTTGC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
151 ATGATTGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
201 CGTCGCCACT TGGACGGGGC AGCCGCCAC GCGGGCGGC GTTCTGGTCG
251 GCTTGACTAT CTTTGGCTG GCTGCGCGGA TTGCCGCCTT TATCCCGGGT
301 TGGGGTGCGT CGGCAAGCGG CATACTCGGT ACGCTGTTTT TCTGGTACGG
10 351 CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TTCGCAGAAT CAACGCAACT
401 ATGTTGCCGT GTTCGCGCTG TTCGCTTGG GCGGCACGCA TGGCGGTTTC
451 CACGTCCAGC TGCACAACGG CAACCTAGGC GGACTCTTGA GCGGATTGCA
501 GTCGGGCTTG GTGATGGTGT CGGGTTTAT CGGTCTGATT GGTACGCGGA
551 TTATTTCTGT TTTTACGTCC AAACGCTTGA ATGTGCCGCA GATTCCCACT
15 601 CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTGCCCATGC TGAATGCCAT
651 GCTGATGGCG CACGGTGTGT TGGCTTGGCT GTCTGCCGTT TTTGCCCTTG
701 CGGCAGGTGT GATTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAACCC
751 GTGTTGAAAG AGCCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAAA CCCGCTTTCC
20 851 TCAATCTGGG TGTGCATCTG ATCGGGGTG GCGGTATCGG CGTGCTGACT
901 TTGGGCATGA TGGCGCGTAC CGCGCTTGGT CACACGGGCA ATCCGATTTA
951 TCCGCCGCCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
1001 CCGCCGTCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATCCGCA CCTCTTCGGT TTGTTTGCA CTCGCGCTTT TGGTGTATGC
25 1101 GTGGAAGTAT ATTCCTTGGC TGATTGCTC GCGTTCGGAC GGCAGGCCCG
1151 GTTGA
  
```

This corresponds to the amino acid sequence <SEQ ID 708; ORF47-1>:

```

      1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WFGYTGTHE LSGFYWHAHE
      51 MIWGYAGLVV IAFLLTAVAT WTGPPTRGV VLVGLTIFWL AARIAAFIPG
30 101 WGASASGILG TLFYWGAVC MALPVIRSON QRNYVAVFAL FVLGGTHAAF
151 HVQLHNGNLG GLLSGLQSGL VMVSGFIGLI GTRIISFPTS KRLNVPQIPS
201 PKWVAQASLW LPMLTAMLMH HGVLAWSAV FAFAGVIFT VQVYRWYKP
251 VLKEPMLWL FAGYLFYGLG LIAVGASYFK PAFNLGVHL IGVGIGVLT
301 LGMMARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
35 351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*
  
```

Computer analysis of this amino acid sequence predicts a leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF47 shows 99.4% identity over a 172aa overlap with an ORF (ORF47a) from strain A of *N.*

40 *meningitidis*:

```

      10      20      30      40      50      60
orf47.pep MKFTKHPVWMAFRPFYSLAALYGALSVLLWFGYTGTHXLSGFYWHAHEMIWGYAGLVV
      10      20      30      40      50      60
45 orf47a  MKFTKHPVWMAFRPFYSLAALYGALSVLLWFGYTGTHLSGFYWHAHEMIWGYAGLVV
      70      80      90      100     110     120
orf47.pep IAFLLTAVATWTGPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGLFFWYGAVC
      70      80      90      100     110     120
50 orf47a  IAFLLTAVATWTGPPTRGVVLVGLTIFWLAARIAAFIPGWGASASGILGLFFWYGAVC
      130     140     150     160     170
orf47.pep MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVM
      130     140     150     160     170     180
55 orf47a  MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVMVSGFIGLI
      130     140     150     160     170     180
60 orf47a  GTRIISFPTSKRLNVPQIPSPKWVAQASLWLPMLTAMLMHGVMPWLSAFAFAAGVIFT
      190     200     210     220     230     240
  
```

The complete length ORF47a nucleotide sequence <SEQ ID 709> is:

```

1  ATGAAATTTA CCAAGCACCC CGTTTGGGCA ATGGCGTTCC GCCCGTTTTA
51  TTCCTGCGCG GCTCTGTACG GCGCATTGTC CGTATTGCTG TGGGGTTTCG
101 GCTACACGGG AACGCACGAG CTGTCCGGTT TCTATTGGCA CGCGCATGAG
151 ATGATTGGGG GTTATGCCGG ACTGGTCGTC ATCGCCTTCC TGCTGACCGC
201 CGTCGCCACT TGGACGGGGC AGCCGCCAC GCGGGCGGGC GTTCTGGTCG
251 GCTTGACTAT CTTTTGGCTG GCTGCGCGGA TTGCCGCTT TATCCCGGT
301 TGGGGTGGCT CGGCAAGCGG CATACTCGGT ACGCTGTTTT TCTGGTACGG
351 CGCGGTGTGC ATGGCTTTGC CCGTTATCCG TTCGCAGAAT CAACGCAATT
401 ATGTTGCCGT GTTCGCGCTG TTCGTCTTGG GCGGTACGCA CGCGGCTTC
451 CACGTCCAGC TGCACAACGG CAACCTAGGC GGA CTCTTGA GCGGATTGCA
501 GTCGGGCTTG GTGATGGTGT CGGGTTTTAT CGGTCTGATT GGTACGCGGA
551 TTATTTCTGT TTTTACGTCC AAACGGTTGA ATGTGCCGCA GATTCCCACT
601 CCGAAATGGG TGGCGCAGGC TTCGCTGTGG CTGCCCATGC TGACCGCAT
15 651 GCTGATGGCG CACGGCGTGA TGCCTTGGCT GTCGGCGGCT TTCGCGTTT
701 CGGCAGGTGT GATTTTTACC GTGCAGGTGT ACCGCTGGTG GTATAAGCCT
751 GTGTTGAAAG AGCCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAAA CCGCTTTCC
851 TCAATCTGGG TGTGCATCTG ATCGGGGTCT GCGGTATCGG CGTGCTGACT
901 TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CATACGGGCA ATCCGATTTA
951 TCCGCCGCC AAAGCCGTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
1001 CCGCGCTCCG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATACGCA CCTCTTCGGT TTTGTTTGCA CTCGCGCTTT TGGTGTATGC
1101 GTGGAAGTAT ATTCCTTGGC TGATTCGTCC GCGTTCGGAC GGCAGGCCCG
1151 GTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 710>:

```

1  MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFGYTGTHE LSGFYWHAHE
51  MIWGYAGLVV IAFLLTAVAT WTGQPPTRGV VLVGLTIFWL AARIAAFIPG
101 WGASASGILG TLFFWYGAVC MALPVIRSON QRNYVAVFAL FVLGGTHAAF
151 HVQLHNGNLG GLLSGLQSGV VMVSGFIGLI GTRIIISFFTS KRLNVQIPSP
201 PKWVAQASLW LPMLTAMLMH HGVMPWLSAA FAFAAGVIFT VQVYRWYKYP
251 VLKEPMLWIL FAGYLFGLG LIAVGASYFK PAFLNLGVHL IGVGIGVLT
301 LGMMARTALG HTGNPIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

```

ORF47a and ORF47-1 show 99.2% identity in 384 aa overlap:

```

10 20 30 40 50 60
orf47a.pep MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV
orf47-1 MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFGYTGTHELSGFYWHAHEMIWGYAGLVV
10 20 30 40 50 60
70 80 90 100 110 120
orf47a.pep IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWASASGILGLTFFWYGAVC
orf47-1 IAFLLTAVATWTGQPPTRGVVLVGLTIFWLAARIAAFIPGWASASGILGLTFFWYGAVC
70 80 90 100 110 120
130 140 150 160 170 180
orf47a.pep MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLSGLQSGVLMVSGFIGLI
orf47-1 MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLSGLQSGVLMVSGFIGLI
130 140 150 160 170 180
190 200 210 220 230 240
orf47a.pep GTRIIISFFTSKRLNVQIPSPKWVAQASLWLPMLTAMLMHGVMPWLSAAFAFAAGVIFT
orf47-1 GTRIIISFFTSKRLNVQIPSPKWVAQASLWLPMLTAMLMHGVMPWLSAAFAFAAGVIFT
190 200 210 220 230 240
250 260 270 280 290 300
orf47a.pep VQVYRWYKYPVLKEPMLWILFAGYLFGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT
orf47-1 VQVYRWYKYPVLKEPMLWILFAGYLFGLGLIAVGASYFKPAFLNLGVHLIGVGIGVLT
250 260 270 280 290 300
310 320 330 340 350 360

```

-406-

5	orf47a.pep	LGMMARTALGHTGNPIYPPPKAVPVAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
	orf47-1	LGMMARTALGHTGNPIYPPPKAVPVAFWLMAATAVRMVAVFSSGTAYTHSIRTSSVLFA
		310 320 330 340 350 360
10	orf47a.pep	LALLVYAWKYIPWLIRPRSDGRPGX
	orf47-1	LALLVYAWKYIPWLIRPRSDGRPGX
		370 380

Homology with a predicted ORF from *N.gonorrhoeae*

ORF47 shows 97.1% identity over 172 aa overlap with a predicted ORF (ORF47ng) from *N.gonorrhoeae*:

15	ORF47	MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV	60
	ORF47ng	MKFTKHPVWAMAFRPFYSLAALYGALSVLLWGFYGTGTHELSGFYWHAHEMIWGYAGLVV	60
20	ORF47	IAFLLTAVATWTGQPPTRGGLVGLTIFWLAARIAAFIPGWGASASGILGTLFFWYGAVC	120
	ORF47ng	IAFLLTAVATWTGQPPTRGGLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAVC	120
25	ORF47	MALPVIRSONQRNYVAVFALFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVM	172
	ORF47ng	MALPVIRSONRRNYVAVFAIFVLGGTHAAFHVQLHNGNLGGLLSGLQSGLVMVWGFIGLI	180

The ORF47ng nucleotide sequence <SEQ ID 711> is predicted to encode a protein comprising amino acid sequence <SEQ ID 712>:

30	1	<u>MKFTKHPVWA</u>	<u>MAFRPFYSLA</u>	ALYGALSVLL	WGFGYTGT	THE	LSGFYWHAHE
	51	<u>MIWGYAGLVV</u>	<u>IAFLLTAVAT</u>	<u>WTGQPPTRG</u>	<u>VLVGLTAFWL</u>	<u>AARIAAFIPG</u>	
	101	<u>WGAAASGILG</u>	<u>TLFFWYGAVC</u>	<u>MALPVIRSON</u>	<u>RRNYVAVFAI</u>	<u>FVLGGTHAA</u>	
	151	<u>HVQLHNGNLG</u>	<u>GLLSGLQSGL</u>	<u>VMVWGFIGLI</u>	<u>GMKIIISFFTS</u>	<u>KRLKLPQIPS</u>	
	201	<u>PKWVAHASLW</u>	<u>LPMLNAILMA</u>	<u>HRVMPWLSAA</u>	<u>FPFAAGVIFT</u>	<u>VQVYAGGITP</u>	
	251	<u>IEETSCGSVA</u>	<u>GICYRLGNSS</u>	G			

The predicted leader peptide and transmembrane domains are identical (except for an Ile/Ala substitution at residue 87 and an Leu/Ile substitution at position 140) to sequences in the meningococcal protein (see also *Pseudomonas stutzeri* orf396, accession number e246540):

40	TM segments in ORF47ng			
	INTEGRAL	Likelihood = -5.63	Transmembrane	52 - 68
	INTEGRAL	Likelihood = -3.88	Transmembrane	169 - 185
	INTEGRAL	Likelihood = -3.08	Transmembrane	82 - 98
	INTEGRAL	Likelihood = -1.91	Transmembrane	134 - 150
	INTEGRAL	Likelihood = -1.44	Transmembrane	107 - 123
	INTEGRAL	Likelihood = -1.38	Transmembrane	227 - 243

Further work revealed the complete gonococcal DNA sequence <SEQ ID 713>:

45	1	ATGAAATTTA	CCAAACATCC	CGTCTGGGCA	ATGGCGTTCC	GCCCGTTTTA
	51	TTCACCTGGC	GCACTGTACG	GCGCATTGTC	CGTATTGCTG	TGGGGTTTCG
	101	GCTACACGGG	AACGCACGAG	CTGTCGGGTT	TCTATTGGCA	CGCGCATGAG
	151	ATGATTGGG	GTTATGCCGG	TCTCGTCGTC	ATCGCCTTCC	TGCTGACCGC
50	201	CGTCGCCACT	TGGACGGGAC	AGCCGCCAC	GAGGGGCGGC	GTTCTGGTCG
	251	GCTTGACCGC	CTTTTGCTG	GCTGCGCGGA	TGCGCGCCTT	TATCCCGGGT
	301	TGGGGTGGCG	CGGCAAGCGG	CATACTCGGT	ACGCTGTTT	TCTGGTACGG
	351	CGCGGTGTGC	ATGGCTTTGC	CGTTATCCG	TtcgCAAAC	CGGCGCAACT
55	401	ATGtcgCCGT	ATTCGCAATA	TTGTGCTGG	GCGGTACGCA	TGCGgcgTTC
	451	CACGtccAgc	tGCACAACGG	CAACCTAGGC	GGACTCTTGA	GCGGATTGCA
	501	GTCGGGCCTG	GTTATGGTGT	CGGGCTTAT	CGGCCTGATT	GGGATGAGGA
	551	TTATTTCTGT	TTTTACGTCC	AAACGGTTGA	ACGTGCCGCA	GATTCCCACT
	601	CCGAAATGGG	TGGCGCAGGC	TTCGCTGTGG	CTACCCATGC	TGACCGCCAT

-407-

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115

651 ACTGATGGCG CACGGCGTGA TGCCTTGGCT GTCGGCGGCT TTCGCGTTTG
701 CGGCGGGCGT GATTITTTACC GTACAGGTGT ACCGCTGGTG GTATAAACCC
751 GTATTGAAAG AACCGATGCT GTGGATTCTG TTTGCCGGCT ATCTGTTTAC
801 CGGATTGGGG CTGATTGCGG TCGGCGCGTC TTATTTCAAA CCTGCCTTCC
851 TCAATCTGGG CGTACATCTG ATCGGGGTCG GCGGTATCGG CGTGCTGACT
901 TTGGGCATGA TGGCGCGTAC CGCGCTCGGT CATACGGGCA ATTCGATTTA
951 TCCGCCGCC AAAGCCGTTT CCGTTGCGTT TTGGCTGATG ATGGCGGCAA
1001 CCGCCGTCGG TATGGTTGCC GTATTTTCTT CCGGCACTGC CTACACGCAC
1051 AGCATCCGCA CGTCTTCGGT TTTGTTTGCA CTCGCGCTGC TGGTGTATGC
1101 GTGGAAATAC ATTCCGTGGC TGATCCGTCC GCGTTCGGAC GGCAGGCCCG
1151 GTTGA

This encodes a protein having amino acid sequence <SEQ ID 714; ORF47ng-1>:

1 MKFTKHPVWA MAFRPFYSLA ALYGALSVLL WGFYGTGTHE LSGFYWHAHE
51 MIWGYAGLVV IAFLLTAVAT WTGQPPTRGV VLVLTAFLW AARIAAFIPG
101 WGAASGILG TLEFFWYAVC MALPVIRSON RRNYVAVFAI FVLGGTHAAF
151 HVQLHNGNLG GLLSGLQSLG VMVSGFIGLI GMRIISFFTS KRLNVPQIPS
201 PKWVAQASLW LPMLTAILMA HGVPWLSAA FAFAGVIFT VQVYRWYKYP
251 VLKEPMLWIL FAGYLTGLG LIAVGASYFK PAFLNLGVHL IGVGIGVLT
301 LGMARTALG HTGNSIYPPP KAVPVAFWLM MAATAVRMVA VFSSGTAYTH
351 SIRTSSVLFA LALLVYAWKY IPWLIRPRSD GRPG*

ORF47ng-1 and ORF47-1 show 97.4% identity in 384 aa overlap:

25
30
35
40
45
50
55
60

orf47-1.pep	10	20	30	40	50	60
orf47ng-1	10	20	30	40	50	60
orf47-1.pep	70	80	90	100	110	120
orf47ng-1	70	80	90	100	110	120
orf47-1.pep	130	140	150	160	170	180
orf47ng-1	130	140	150	160	170	180
orf47-1.pep	190	200	210	220	230	240
orf47ng-1	190	200	210	220	230	240
orf47-1.pep	250	260	270	280	290	300
orf47ng-1	250	260	270	280	290	300
orf47-1.pep	310	320	330	340	350	360
orf47ng-1	310	320	330	340	350	360
orf47-1.pep	370	380				
orf47ng-1	370	380				

Furthermore, ORF47ng-1 shows significant homology to an ORF from *Pseudomonas stutzeri*:

65 gnl|PID|e246540 (273914) ORF396 protein [*Pseudomonas stutzeri*] Length = 396
Score = 155 bits (389), Expect = 5e-37

Identities = 121/391 (30%), Positives = 169/391 (42%), Gaps = 21/391 (5%)

```

Query: 7  PVWAMAFRPFYSLAALYGALSVLLWGFYTGTHELSGFY-----WHAHEMIWGYAGLV 59
          P+W +AFRPF+ +LY L++ LW +TG GF WH HEM++G+A +
5  Sbjct: 14 PIWRLAFRPFFLAGSLYALLAIPWVAAWTGLWP--GFQPTGGWLAWHRHEMLFGFAMAI 71

Query: 60 VIAFLLTAVATWTGQPPTRGVVLVGLTAFWLAARIAAFIPGWGAAASGILGTLFFWYGAV 119
          V FLLTAV TWTGQ G LVGL A WLAAR+ ++ G AA L LF
10 Sbjct: 72 VAGFLLTAVQWTGTQAPSGNRLVGLAAVWLAARL-GWLFGLPAAWLAPDLLLFLVALVW 130

Query: 120 CMALPVIRSONRRNYVAVFAIFVLGGTHAAFXXXXXXXXXXXXXXXXXXXXXMVSGFIGL 179
          MA + + +RNY V + ++ G +V+ + L
10 Sbjct: 131 MMAQMLWAVRQKRNPYIVVVLISMLGADVLIILTGLLQGNDAQRQGVLAGLWLVAALMAL 190

Query: 180 IGMRIISFFTSKRLNVPQIPSP-KWVAQASLWLPMLTAILMAHGV----MPWLSAAFAFA 234
          IG R+I FFT + L P W+ A L + A+L A GV P L F A
15 Sbjct: 191 IGGRVIPFFTQRGLGKVDKVPWVWLDVALLVGTGVIALHAFGVAMRPQPLLGLLFV-A 249

Query: 235 AGVIFTVQVYRWYKPVLEPMLWILFAGYLTGLGLIAGVASYF-KPAFXXXXXXXXXXX 293
          GV +++ RW+ K + K +LW L L+ + + +F A
20 Sbjct: 250 IGVGHLLRLMRWYDKGIWKVGLLWSLHVAMLWLVAAFGLALWHFGLLAQSSPSLHALSV 309

Query: 294 XXXXXXXXXXXMMARTALGHTGNSIYPPPKAVPFAFWLXXXXXXXXXXXXFSSGTAYTHSIR 353
          M+AR LGHTG + P + AF L F S +
25 Sbjct: 310 GSMSGLILAMIAVTLGHTGRPLQLPAGIIG-AFVL---FNLGTAARVFLSVAPVPGGLW 365

Query: 354 TSSVLFALALLVYAWKYIPWLIRPRSDGRPG 384
          ++V + LA +Y W+Y P L+ R DG PG
30 Sbjct: 366 LAAVCWTLAFALYVWRYAPMLVAARVDGHPG 396

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 85

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 715>:

```

35      1  ..ATGCCGCTCTG AAGGTTTCAGA CGGCmTCGGT GycGGGGAay CAGAAGyGGT
          51  AGCGCATGCC CAATGAGACT TCGTGGGTTT TGAAGCGGGT GTTTTCCAAG
          101  CGTCCCCAGT TGTGGTAACG GTATCCGGTG TCyAArGTCA GCTTGGGyGT
          151  GATGTCGAAa CCGACACCGG CGATGACACC AAGACCyAmG CTGCTGATrC
          201  TGTkGCTTTT GTGATAGGsA GGTtTGyTGG kmksAsyTTG TAYrATwkkG
40      251  CCTssCwsTG kAGmGCCkTk CkyTGGTkkA swGrwArTAG TCGTGGTtTy
          301  TkTtTyCACc GAATGAACyT GATGTTTAAC GTGTCCGTAG GCGACGCGCG
          351  CGCCGATATA GGGTTTGAAT TTATCGTTGA GTTTGAAATC GTAAATGGCG
          401  GACAAGCCGA GAGAAGAAAC GGCGTGGAAG CTGCCGTTTC CCTGATGTTT
          451  TGTTTGGGTT TCTTTGTAGT TGTTGTTTAT CTCTTCAGTA ACTTTTtTAG
45      501  TAGAAGAATT ACTTTCTTTC CATTtTCTGT AACTGGCATA ATCTGCCGCT
          551  ATTCTCCAGC CGCCGAAATC ..

```

This corresponds to the amino acid sequence <SEQ ID 716; ORF67>:

```

50      1  ..MPSEGSDGXG XGEXEXVAHA QXDFVGFEAG VFQASPVVVT VSGVXXQLGX
          51  DVETDTGDDT KTXAADXVAF VIGRFxGXXL YXXAXXXAX XWXXXSRGF
          101  XXHRMNLmFN VSVGDARADI GFEFIVEFEI VNGGQAERN GVEAAVSLMF
          151  CLGFFVVVVY LFSNFFSRRI TFFPFSVTGI ICRYSPAAEI ..

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.gonorrhoeae*

ORF67 shows 51.8% identity over 199 aa overlap with a predicted ORF (ORF67ng) from

55 *N.gonorrhoeae*:

-409-

	orf67.pep		MPSEGS	DXGXGE	EXVAHQ	DFVGFEAG	30
	orf67ng	TNFEIAVLSGMTVRVFCARPAPVNGGRLKMPSEGS	DGIGIG	ESEAVA	HAQRGF	VGFEAG	146
		90 100 110 120 130 140					
5	orf67.pep	VFAQSPVVVTVSGVXXQLGXDVETDTGDDTKTXAADXVAFVIGRFXGXLYXXXAX					90
	orf67ng	VFAQSPVVAVAGVQGGQAGRDVYAHARHRAEAQAAA	AVFLIGV	FLRMSVR	INRNC	CVSI	206
10	orf67.pep	XWXXXSRGFXHRMNLMFNVSVGDARADIGFEFIVEFEIVNGGQAERRNGVEAAVSLMF					150
	orf67ng	TRVGGKSTCYFFSRIDAVSDVSVGDARTDIGFEFVVEFEIVNGGQAERRNGVECAVFLMF					266
15	orf67.pep	CLGFFVV-----VVYLFNSNFFSRITFF-PFSVTGIICRYSPAAEI					190
	orf67ng	RLLVFYVKLVAAKSFIILSFQLFYVHGIFIVVPFVPTGIIRGDAPAAEVVADRH	PGVDGM				326

The ORF67ng nucleotide sequence <SEQ ID 717> is predicted to encode a protein comprising amino acid sequence <SEQ ID 718>:

	1	MPSETVGSIV	NVGVD	ESVGF	SPPFPSIQHF	YRFHRIHRIR	LFRPPGPMQL
20	51	NRHSHSGNL	GRGVW	ATVLS	DKFPCGQVRI	PACAGMTNFE	IAVLSGMTVR
	101	VFYCARPAPV	NGGRLK	MPSE	GSDGIGIGES	EAVAHAQRGF	VGFEAGVFQA
	151	SPVVAVAGV	QGGQAG	RDVYA	HARHRAEAQA	AAAVFLIGV	FLRMSVRINR
	201	NCCVSITRVG	GKSTCY	FFSR	IDAVSDVSVG	DARTDIGFEF	VVEFEIVNGG
25	251	QAERRNGVEC	AVFLMF	RLLV	FYVKLVAAKS	FIILSFQLFY	VHGIFIVVPF
	301	PVTGIIRGDA	PAAEVV	ADRH	PGVDGMRTDV	SEIIAYRAYF	VFAWSGWFR
	351	IVGNAFGVG	*				

Based on the presence of a several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

30 Example 86

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 719>

	1	ATGTTTGCTT	TTTAGAAGC	CTTTTTGTC	GAATACGGTT	ATGCGGCTGT
	51	TTTTTTTGT	TTGGTCATCT	GCGGTTTCGG	CGTGCCGATT	CCCGAGGATT
35	101	TGACCTTGGT	AACAGGCGGC	GTGATTTCGG	GTATGGGTTA	TACCAATCCG
	151	CATATTATGT	TTGCAGTCGG	TATGCTCGGC	GTATTGGTCG	GGGACGGCAT
	201	CATGTTCCGC	GCCGGACGAA	TTTGGGGGCA	GAAATCCTA	AGGTTCAAAC
	251	CTATTGCGCG	CATCATGACG	CCGAAACGTT	ATGAGCAGGT	TCAGGAAAAA
	301	TTGAGCAAAT	ACGGTAACTG	GGTCTTATT	GTCGCCCCGT	TCCTGCCCCG
40	351	TTTGAGAACG	GCCGTATTG	TTACAGCCCG	TATCAGCCCG	AAGGTTTCAT
	401	ACTTGCGTTT	TATCATTATG	GATGGACTGG	CCGCA...	

This corresponds to the amino acid sequence <SEQ ID 720; ORF78>:

	1	MFAFLEAFFV	EYGYAAVFFV	LVICGFGVPI	PEDLT	LVTGG	VISGMGYTNP
	51	HIMFAVGMLG	VLVGDGIMFA	AGRIWQXXL	XFXPIAXIMT	PXRYEQVQEK	
	101	FDKYGNWVLF	VARFLPGLRT	AVFVTAGISR	KVSYLRFIIM	DGLAA...	

45 Further work revealed the complete nucleotide sequence <SEQ ID 721>:

	1	ATGTTTGCTT	TTTAGAAGC	CTTTTTGTC	GAATACGGTT	ATGCGGCTGT
	51	TTTTTTTGT	TTGGTCATCT	GCGGTTTCGG	CGTGCCGATT	CCCGAGGATT
	101	TGACCTTGGT	AACAGGCGGC	GTGATTTCGG	GTATGGGTTA	TACCAATCCG
	151	CATATTATGT	TTGCAGTCGG	TATGCTCGGC	GTATTGGTCG	GGGACGGCAT
50	201	CATGTTCCGC	GCCGGACGAA	TTTGGGGGCA	GAAAATCCTA	AGGTTCAAAC
	251	CTATTGCGCG	CATCATGACG	CCGAAACGTT	ATGAGCAGGT	TCAGGAAAAA
	301	TTGAGCAAAT	ACGGTAACTG	GGTCTTATT	GTCGCCCCGT	TCCTGCCCCG
	351	TTTGAGAACG	GCCGTATTG	TTACAGCCCG	TATCAGCCCG	AAGGTTTCAT
	401	ACTTGCGTTT	TATCATTATG	GATGGACTGG	CCGCACTGAT	TTCCGTCCTT
55	451	ATTGCGATT	ATCTGGGCGA	ATACGGTGCG	CACAACATCG	ATTGGCTGAT


```

501 GGCAGAAATG CACAGCCTGC AATCGGGTAT TTTTGTATC TTGGGTATAG
551 GTGCGACCGT TGTCGCTTGG ATTTGGTGGA AAAACGCCA ACGTATCCAG
601 TTTTACCGCA GCAAATTGAA AGAAAAGCGG GCGCAACGCA AAGCCGCCAA
651 GGCAGCCAAA AAAGCCGCGC AAAGCAAACA ATAA

```

5 This corresponds to the amino acid sequence <SEQ ID 722; ORF78-1>:

```

1 MFAFLEAFFV EYGYAAVFFV LVICGFGVPI PEDLTLVTGG VISGMGYTNP
51 HIMFAVGMLG VLVGDGIMFA AGRIWGQKIL REKPIARIMT PKRYEQVQEK
101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRFIIM DGLAALISVP
151 IWIYLGEYGA HNIDWLMAKM HSLSQGI FVI LGIGATVVAW IWWKKRQRIQ
201 FYRSKLKEKR AQRKAACA KAAQSKQ*

```

Computer analysis of this amino acid sequence predicts several transmembrane domains, and also gave the following results:

Homology with the dedA homologue of *H. influenzae* (accession number P45280)

ORF78 and the dedA homologue show 58% aa identity in 144aa overlap:

```

15 Orf78: 4 FLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGM--GYTNPHIMFAVGMLGV 61
    FL FF EYGY AV FVL+ICGFGVPIPED+TLV+GGVI+G+ N H+M V M+GV
    DedA: 20 FLIGFFTEYGYWAVLFLVLIICGFGVPIPEDITLVSGGVIAGLYPENVNSHLMLLVSMIGV 79

20 Orf78: 62 LVGDGIMFAAGRIWGQXXLXFXPIAXIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRTA 121
    L GD M+ GRI+G L F PI I+T R V+EKF +YGN VLFVARFLPGLR
    DedA: 80 LAGDSCHYWLGRYGTIKILRFRPIRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAP 139

    Orf78: 122 VFVTAGISRKVSYLRFIIMDGLAA 145
    +++ +GI+R+VS+Y+RF+++D AA
25 DedA: 140 IYVSGITRRVSYRVFLIDFCAA 163

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

ORF78 shows 93.8% identity over a 145aa overlap with an ORF (ORF78a) from strain A of *N. meningitidis*:

```

30 orf78.pep      10      20      30      40      50      60
    MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG
    orf78a        10      20      30      40      50      60
    MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTLVTGGVISGMGYTNPHIMFAVGMLG

35 orf78.pep      70      80      90      100     110     120
    VLVGDGIMFAAGRIWGQXXLXFXPIAXIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT
    orf78a        70      80      90      100     110     120
    VLVGDGIMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT

40 orf78.pep      130     140
    AVFVTAGISRKVSYLRFIIMDGLAA
    orf78a        130     140
    AVFVTAGISRKVSYLRFIIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLSQGI FIA

45 orf78a        130     140     150     160     170     180

```

The complete length ORF78a nucleotide sequence <SEQ ID 723> is:

```

1 ATGTTTGCCC TTTTGAAGC CTTTTTGTG GAATACGGCT ATGCGGCCGT
51 GTTTTTCGTT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAGGATT
101 TGACCTTGGT AACAGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG
151 CATATTATGT TTGCAGTCGG TATGCTCGGC GTATTGGTCG GGGACGGCAT
201 CATGTTGCGC GCCGACGCA TCTGGGGGCA GAAATCCTC AAGTTCAAAC
251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCACAGGT TCAGGAAAAA
301 TTCACAAAT ACGGCAACTG GGTGTTATT GTCGCTCGTT TCCTGCCCGG
351 TTTGCGGACT GCCGTTTTCG TTACCGCCGG CATCAGCCGC AAAGTATCGT
55 ATCTGCGCTT TCTGATTATG GACGGGCTTG CCGCGCTGAT TTCCGTGCCC
401 ATCTGCGCTT TCTGATTATG GACGGGCTTG CCGCGCTGAT TTCCGTGCCC
451 GTTTGGATT ACTTGGGCGA GTACGGCGCG CACAACATCG ATTGGCTGAT

```

-411-

501 GGCAGAAATG CACAGCCTGC AATCCGGCAT CTTTCATCGCA TTGGGCGTGC
 551 TGGCGGCGGC GCTGGCGTGG TTCTGGTGGC GCAAACGCCG ACATTATCAG
 601 CTTTACCGCG CACAATTGAG CGAAAAACGC GCCAAACGCA AGGCGGAAAA
 651 GGCAGCGAAA AAAGCGGCAC AGAAGCAGCA GTAA

5 This encodes a protein having amino acid sequence <SEQ ID 724>:

1 MFALLEAFFV EYGAAVFFV LVICGFGVPI PEDLTlVTGG VISGMGYTNP
 51 HIMFAVGMLG VLVGDGIMFA AGRIWGQKIL KFKPIARIMT PKRYAQVQEK
 101 FDKYGNWVLF VARFLPGLRT AVFVTAGISR KVSYLRLIM DGLAALISVP
 151 VWIYLGEYGA HNIDWLMAMK HSLQSGIFIA LGVLAALAW FWRKRHHYQ
 201 LYRAQLSEKR AKRKA EKA KAAQKQ*

ORF78a and ORF78-1 show 89.0% identity in 227 aa overlap:

		10	20	30	40	50	60
	orf78a.pep	MFALLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTlVTGGVISGMGYTNP	HIMFAVGMLG				
15	orf78-1	MFAFLEAFFVEYGYAAVFFVLVICGFGVPIPEDLTlVTGGVISGMGYTNP	HIMFAVGMLG				
		10	20	30	40	50	60
	orf78a.pep	VLVGDGIMFAAGRIWGQKILKFKPIARIMTPKRYAQVQEKFDKYGNWVLFVARFLPGLRT					
20	orf78-1	VLVGDGIMFAAGRIWGQKILKFKPIARIMTPKRYEQVQEKFDKYGNWVLFVARFLPGLRT					
		70	80	90	100	110	120
	orf78a.pep	AVFVTAGISRKVSYLRLIMDGLAALISVPVWIYLGEYGAHNIDWLMAMKHSLQSGIFIA					
25	orf78-1	AVFVTAGISRKVSYLRFIIMDGLAALISVPWIYLGEYGAHNIDWLMAMKHSLQSGIFVI					
		130	140	150	160	170	180
	orf78a.pep	LGVLAAALAWFWRKRHHYQLYRAQLSEKRAKRAEKAAKAAQKQX					
30	orf78-1	LGIGATVVAWVWKKRQRIQFYRSKLKEKRAQKAAKAAQSKQX					
		190	200	210	220		
35							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF78 shows 97.4% identity over 38 aa overlap with a predicted ORF (ORF78ng) from *N. gonorrhoeae*:

40	orf78.pep	XXLXFXPIAXIMTPXRYEQVQEKFDKYGNWVLFVARFLPGLRTAVFVTAGISRKVSYLRF	137
	orf78ng	YPVLFVARFLPGLRTAVFVTAGISRKVSYLRF	32
	orf78.pep	IIMDGLAA	145
45	orf78ng	LIMDGLAALISVPVWIYLGEYGAHNIDWLMAMKHSLQSGIFIALGVLAALAWFWRKR	92

The ORF78ng nucleotide sequence <SEQ ID 725> is predicted to encode a protein comprising amino acid sequence <SEQ ID 726>:

1 ..YPVLFVARFL PGLRTAVFVT AGISRKVSYL RFLIMDGLAA LISVPVWIYL
 51 GEYGAHNIDW LMAKMHSLQSGIFIALGVLA AALAWFWRKR RHHYQLYRAQ
 101 LSEKRAKRA EKAAKAAQK QQ*

Further work revealed the complete gonococcal nucleotide sequence <SEQ ID 727>:

1 atgtttgccc tttTggaagc CTTTTTTGTC GAAtacggCt atgcGGCCGT
 51 GTTTTTCGTT TTGGTCATCT GCGGTTTCGG CGTGCCGATT CCCGAAGATT
 101 TGACCTTGGT AACGGGCGGC GTGATTTCGG GTATGGGTTA TACCAATCCG
 151 CATATTATGT TTGCGGTCGG TATGCTCGGC GTGTTGGCGG GCGACGGCGT
 201 GATGTTTGCC GCCGGACGCA TCTGGGGGCA GAAATCCTC AAGTTCAAAC
 251 CGATTGCGCG CATCATGACG CCGAAACGTT ACGCGCAGGT TCAGGAAAAA
 301 TTCGACAAAT ACGGCAACTG GGTTCGTTT GTCGCCCGTT TCCTGCCGGG

This corresponds to the amino acid sequence <SEQ ID 728; ORF78ng-1>:

10

ORF78ng-1 and ORF78-1 show 88.1% identity in 227 aa overlap:

15

20

25

30

35

Furthermore, orf78ng-1 shows homology to the dedA protein from *H.influenzae*:

40

```

sp|P45280|YG29_HAEIN HYPOTHETICAL PROTEIN HI1629 >gi|1073983|pir||D64133 dedA
protein (dedA) homolog - Haemophilus influenzae (strain Rd KW20)
>gi|1574476 (U32836) dedA protein (dedA) [Haemophilus influenzae] Length = 212
Score = 223 bits (563), Expect = 7e-58
Identities = 108/182 (59%), Positives = 140/182 (76%), Gaps = 2/182 (1%)

```

45

Query: 5 LEAFFVEYGYAAVFFVLVICGFGVIPEDLTLTGGVISGM--GYTNPHIMFAVGMLGVL 62
L FF EYGY AV FVL+ICGFGVIPED+TLV+GGVI+G+ N H+M V M+GVL
Sbjct: 21 LIGFFTEYGYWAVLFVLIICGFGVIPEDITLVSGGVIAGLYPENVNSHLMLLVSMIGVL 80

50

Query: 63 AGDGVMFAGAGRIWGQKILKFKPIARIMTPKRYAQVQEKFKDYGNWLVFVARFLPGLRTAV 122
AGD M+ GRI+G KIL+F+PI RI+T +R V+EKF +YGN VLFVARFLPGLR +
Sbjct: 81 AGDSCMYWLGRIGYTKILRFRPIRRIVTLQRLRMVREKFSQYGNRVLFVARFLPGLRAPI 140

55

```
Query: 123 FVTAGISRKVSYLRFLIMDGLAALISVPVWIYLGEYGAHNIDWLMAKMHSLSQSGIFIALG 182
      ++ +GI+R+VSY+RF+++D AA+ISVP+WIIYLGE GA N+DWL ++ Q I+I +G
Sbjct: 141 YMVSGITRRVSYVRFLVIDFCAAIIISVPWIWIIYLGEIGAKNLDWLHTQIQKGQIVIIYIFIG 200

Query: 183 VL 184
      L
Sbjct: 201 YL 202
```

5 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 729>:

15 This corresponds to the amino acid sequence <SEQ ID 730; ORF79>:

Further work revealed the complete nucleotide sequence <SEQ ID 731>:

30 This corresponds to the amino acid sequence <SEQ ID 732; ORF79-1>:

35 Computer analysis of this amino acid sequence revealed a putative leader peptide and also gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF79 shows 94.6% identity over a 147aa overlap with an ORF (ORF79a) from strain A of *N. meningitidis*:

BNSDOCID <WO__892457BA2_1_>

```

              130      140
orf79.pep    VTLKFKNAKAQTVQLEVKIAPMPAMNH
              |||||
5  orf79a     VTLKFKNAKAQTVQLEVKTAPMSAMDHGHHHGEAHQH
              130      140      150

```

The complete length ORF79a nucleotide sequence <SEQ ID 733> is:

```

1  ATGAAANAAC TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
51 TTCCGCCGCC GGAATCCACG TTGAGGACGG CTGGGCGCGC ACCACCGTCG
101 AAGGTATGAA AATGGGCGGC GCGTTCATGA AAATCCACAA CGACGAAGCC
151 AAACAAGACT TTTTGCTCGG CGGAAGCAGC CCTGTTGCCG ACCGCGTCGA
201 AGTGCATACC CATATCAATG ATAACGGTGT GATGCGGATG CGCGAAGTCG
251 AAGGCGGCGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCCGGC
301 AGCTATCATG TCATGTTTAT GGGTNTGAAA AAACAATTAA AAGANGGCCA
15 351 CAAGATTCCC GTTACCCTGA AATTAAAAA CGCCAAAGCA CAAACCGTCC
401 AACTGGAAGT CAAACCCGGC CCGATGTCGG CAATGGACCA CGGTCATCAC
451 CACGGCGAAG CGCATCAGCA CTAA

```

This encodes a protein having amino acid sequence <SEQ ID 734>:

```

1  MKXLLAAVMM AGLAGAVSAA GIHVEDGWAR TTVEGMKMG GAFMKIHNDEA
20 51 KQDFLLGGSS PVADRVEVHT HINDNGVMRM REVEGGVPLE AKSVTELKPG
101 SYHVMFMGXK KQLKXGDKIP VTLKFKNAKA QTVQLEVKTA PMSAMDHGHH
151 HGEAHQH*

```

ORF79a and ORF79-1 show 94.9% identity in 157 aa overlap:

```

25      10      20      30      40      50      60
orf79a.pep MKXLLAAVMMAGLAGAVSAAGIHVEDGWAR TTVEGMKMG GAFMKIHNDEAKQDFLLGGSS
      || |||||
orf79-1     MKXLLAAVMMAGLAGAVSAAGVHVEDGWAR TTVEGMKIG GAFMKIHNDEAKQDFLLGGSS
      10      20      30      40      50      60

30      70      80      90      100     110     120
orf79a.pep PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGXKKQLKXGDKIP
      |||||
orf79-1     PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
      70      80      90      100     110     120

35      130     140     150
orf79a.pep VTLKFKNAKAQTVQLEVKTAPMSAMDHGHHHGEAHQH
      |||||
orf79-1     VTLKFKNAKAQTVQLEVKIAPMPAMNHGHHHGEAHQH
      130     140     150

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF79 shows 96.1% identity over 76 aa overlap with a predicted ORF (ORF79ng) from *N.gonorrhoeae*:

```

45      orf79.pep    FMKIHNDEAKQDFLLGGSSPVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGS 101
      orf79ng              INDNGVMRMREVKGVPLEAKSVTELKPGS 30

50      orf79.pep    YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKIAPMPAMNH 147
      orf79ng              YHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEVKTAPMSAMNHGHHHGEAHQH 86

```

An ORF79ng nucleotide sequence <SEQ ID 735> was predicted to encode a protein comprising amino acid sequence <SEQ ID 736>:

```

55      1  ..INDNGVMRM EVKGGVPLEA KSVTELKPGS YHVMFMGLKK QLKEGDKIPV
      51  TLKFKNAKAQ TVQLEVKTAP MSAMNHGHHH GEAHQH*

```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 737>:

```

      1 ATGAAAAAAT TATTGGCAGC CGTGATGATG GCAGGTTTGG CAGGCGCGGT
    51 TTccgccgCc GGagTccAtG TCGaggACGG CTGGGCGCGc accaCTGtcg
   101 aaggtATgaa aatggGCGGC GCgttCATga aaATCCACAA CGACGaaGcc
   151 atacaaGACT ttgtgcTCgg CGGaagcatg cccgttgccg accgcGTCGA
    201 AGTGCAtaca cacATCAACG ACAACGGCGT GATGCGTATG CGCGAAGTCA
    251 AAGGCGGCGT GCCTTTGGAG GCGAAATCCG TTACCGAACT CAAACCCGGC
    301 AGCTATCACG TGATGTTTAT GGGTTTGAAA AAACAACTGA AAGAGGGCGA
    351 CAAGATTCCC GTTACCCTGA AATTAAAAA CGCCAAAGCG CAAACCGTCC
   401 AACTGGAAGT CAAACCCGCG CCGATGTCGG CAATGAACCA CGGTCATCAC
   451 CACGGCGAAG CGCATCAGCA CTAA

```

This corresponds to the amino acid sequence <SEQ ID 738; ORF79ng-1>:

```

      1 MKKLLAAVMM AGLAGAVSAA GVHVEDGWAR TTVEGMKMGG AFMKIHND EA
    51 IQDFVLGGSM PVADRVEVHT HINDNGVMRM REVKGGVPLE AKSVTELKPG
   101 SYHVMFMGLK KQLKEGDKIP VTLKFKNAKA QTVQLEVKTA PMSAMNHGHH
   151 HGEAHQH*

```

ORF79ng-1 and ORF79-1 show 95.5% identity in 157 aa overlap:

```

      10      20      30      40      50      60
orf79-1.pep MKKLLAAVMMAGLAGAVSAAAGVHVEDGWAR TTVEGMKIGGAFMKIHND EAKQDFLGGSS
      10      20      30      40      50      60
orf79ng-1   MKKLLAAVMMAGLAGAVSAAAGVHVEDGWAR TTVEGMKMGGA FMKIHND EAIQDFVLGGSM

      70      80      90     100     110     120
orf79-1.pep PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP
      70      80      90     100     110     120
orf79ng-1   PVADRVEVHTHINDNGVMRMREVEGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIP

      130     140     150
orf79-1.pep VTLKFKNAKAQTVQLEVKIAPMPAMNHGHHHGEAHQH*
      130     140     150
orf79ng-1   VTLKFKNAKAQTVQLEVKTAPMSAMNHGHHHGEAHQH*

```

Furthermore, ORF79ng-1 shows significant homology to a protein from *Aquifex aeolicus*:

```

35 gi|2983695 (AE000731) putative protein [Aquifex aeolicus] Length = 151
   Score = 63.6 bits (152), Expect = 6e-10
   Identities = 38/114 (33%), Positives = 58/114 (50%), Gaps = 1/114 (0%)

40 Query: 24 VEDGWAR TTVEGMKMGGA FMKIHND EAIQDFVLGGSM PVADRVEVHTHINDNGVMRMREV 83
      V+  W      G      M I N+  D++G  +A RVE+H  + +N V +M
      Sbjct: 27 VKHPVMEPPPGPN TMMGM IIVNEGDEPDY LIGAKTDIAQRVELHKT V IENDVAKMVPQ 86

      Query: 84 KGGVPLEAKSVTELKPGSYHVMFMGLKKQLKEGDKIPVTLKFKNAKAQTVQLEV 137
      + + + K  E K  YHVM +GLKK++KEGDK+ V L F+ +  TV+  V
45      Sbjct: 87 ER-IEIPPKGKVEFKHHGYHVM IIGLKKRIKEGDKVKVELIFEKSGKITVEAPV 139

```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF79-1 (15.6kDa) was cloned in the pET vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 18A shows the results of affinity purification of the His-fusion protein. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 18B) These experiments confirm that ORF79-1 is a surface-exposed protein, and that it is a useful immunogen.

	1	ATGACGGTAA	CTGCGGCCGA	AGGCGGCCAA	GCTGCCAAGG	CGTTAAAAAA
5	51	ATATCTGATT	ACGGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATCAGCTCGT	CAACCTGGCT
	151	CCGAAGCAAT	GGCGGCGCGA	ATATGTTTTG	TGGTTTTAATA	TCCCGGGGCT
	201	GGGCGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTGTTTGCCG
	251	CCAACGTATT	GGGTGCGCAG	ATCCTCGCCG	CGTGGGACAG	CCTGTTGGGG
10	301	CGGATTCCGG	TTGTGAAATc	CATCTATTCC	AGTGTGAAAA	AAGTATCCGA
	351	ATAcgTGCTG	TTCGACAGCA	GCCGTTTCGT	TAAACGCCG	GTACTIONG
	401	CGTTTCCCCA	GCCCCGGTATT	TGGACGATyG	CTTTCTGTGC	AGGGCGAGTG
	451	TCGAATGCGG	TTAAGGCCGC	ATTGCCGAAs	CAGGGCGATT	ATCTTTCCGT
	501	GTATGTTCCG	ACCACGCCGA	ATCCGACCCG	CGGTTACTAT	ATTATGGTAA
15	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AsCATTGAAA
	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAAC
	651	ATTGGCAsGA	CCTATGCCGT	CTGAAAAGGC	GGATTTGCCC	GAACAACAAT
	701	AA				

20

1	MTVTAAEGGK	AAKALKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNL
51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFVTG	LFAANVLGRQ	ILAAWDSLLG
101	RIPVVKSIYS	SVKKVSEYVL	SDSSRSFKTP	VLVPFPQPGI	WTIAFVSGQV
151	SNAVKAALPX	DGDYLSVYVP	TFNPTGGY	IMVKSDVRE	LDMSVDXLK
201	YVISLGMVIP	DDLPKTKLAX	PMPSEKADLP	EQQ*	

	1	ATGACGGAAC	nTGC GGCCGA	AGGCGGCAAA	GCTGCCAArG	CGTTAAAAAA
	51	ATATCTGATT	ACGGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATCAGCTCGT	CAACCTGCTG
	151	CCGAAGCAAT	GGCGGGCGCA	ATATGTTTTG	GGGTTTTAATA	TCCCGGGGCT
30	201	GGGCGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TGTTTTGCCG
	251	CCAACGTATT	GGTCTGGCAG	ATCCTCGCCG	CGTGGGACAG	CCTGTTGGGG
	301	CGGATTCCGG	TTGTGAAATC	CATCTATTCT	AGTGTGAAAA	AAGTATCCGA
	351	ATCGCTGCTG	TCCGACAGCA	GCCGTTTCGT	TAAACAGCCG	GTACTCGTGC
	401	CGTTTTCCCA	GCCCCGTATT	TGGACGATTG	CTTTCGTGTC	AGGGCAGGTG
35	451	TCGAATGCGG	TTAAGGCCGC	ATTGCCGAAG	GACGGCGATT	ATCTTTCCGT
	501	GTATGTTCCG	ACCACGCCGA	ATCCGACCGG	CGGTTACTAT	TATTATGTAA
	551	AGAAAAAGCA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCATTGAAA
	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAC
	651	ATTGGCAGGA	CCTATGCCGT	CTGAAAAGGC	GGATTTGCCC	GAACAACAAT
40	701	AA				

45

1	MTEXXAEGGK	AAKALKKYLI	TGILVWLP	IA VTVWVVS	YIV SASDQLV	NLL
51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFVTG	LFAANVLGRQ	ILAAWDS	LLG
101	RIPVVKSIYS	SVKKVSESLL	SDSSRSFKTP	VLVPFPQPGI	WTIAFVS	GOV
151	SNVAKAALPK	DDYLSVYVP	TPNPNTGGYI	IMVKKSDVRE	LDMSVDE	ALK
201	YVISLGMVIP	DGLPVKTLAG	PMPSEKADLP	EQQ*		

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 *meningitidis:*

ENSDOCID: <WO -9924578A2_1_>

		70	80	90	100	110	120
	orf98.pep	GFNIPGLGVIVAIAVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKVSEYVL					
5	orf98a	GFNIPGLGVIVAIAVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKVXSLL					
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf98.pep	SDSSRSFKTPVLVPFPQPGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY					
10	orf98a	SDSSRSFKTPVLVPFPQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY					
		130	140	150	160	170	180
		190	200	210	220	230	
15	orf98.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDL PVKTLAXPMPSEKADLPEQQX					
	orf98a	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDL PVKTLAXPMPSEKADLPEQQX					
		190	200	210	220	230	

The complete length ORF98a nucleotide sequence <SEQ ID 743> is:

20	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA
	51	ATATCTGATT	ACGGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT
	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ATCAGCTCGT	CAACCTGCTG
	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTTAATA	TCCCGGGGCT
25	201	GGGCGTTATC	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTTGCCG
	251	CAAACGTATT	GGGCCGGCAG	ATTCTTGCCG	CGTGGGACAG	CTGTGTGGGG
	301	CGGATTCCGG	TTGTGAAGTC	CATCTATTCG	AGTGTGAAAA	AAGTATCCGA
	351	NTCGTTGCTG	TCCGACAGCA	GCCGTTCTGT	TAAAACACCA	GTACTCGTGC
	401	CGTTTCCCCA	ATCGGGTATT	TGGACAATCG	CATTCTGTGC	CGGTCAAGTG
	451	TCGAATGCGG	TTAAGGCCGC	ATTGCCGAAG	GACGGCGATT	ATCTTTCCGT
30	501	GTATGTTCCG	ACCACGCCGA	ATCCGACCGG	CGGTTACTAT	ATTATGGTAA
	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA
	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAC
	651	ATTGGCAGGA	CCTATGCCGT	CTGAAAAGGC	GGATTGCCCC	GAACAACAAT
	701	AA				

35 This encodes a protein having amino acid sequence <SEQ ID 744>:

	1	MTEPAAEGGK	AAKALKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNLL
	51	PKQWRPQYVL	GFNIPGLGVI	VIAVLFVTG	LFAANVLGRQ	ILAAWDSLLG
	101	RIPVVKSIYS	SVKKVXSLL	SDSSRSFKTP	VLVPFPQSGI	WTIAFVSGQV
	151	SNAVKAALPK	DGDYLSVYVP	TPNPTGGYY	IMVKKSDVRE	LDMSVDEALK
40	201	YVISLGMVIP	DDL PVKTLAG	PMPSEKADLP	EQQ*	

ORF98a and ORF98-1 show 98.7% identity in 233 aa overlap:

		10	20	30	40	50	60
	orf98a.pep	MTEPAAEGGKAAKALKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNLL	PKQWRPQYVL	
45	orf98-1	MTEPAAEGGKAAKALKKYLI	TGILVWLPIA	VTWVVSIV	SASDQLVNLL	PKQWRPQYVL	
		10	20	30	40	50	60
		70	80	90	100	110	120
	orf98a.pep	GFNIPGLGVIVAIAVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKVXSLL					
50	orf98-1	GFNIPGLGVIVAIAVLFVTGLFAANVLGRQILAAWDSLLGRIPVVKSIYSSVKVSESL					
		70	80	90	100	110	120
		130	140	150	160	170	180
	orf98a.pep	SDSSRSFKTPVLVPFPQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY					
55	orf98-1	SDSSRSFKTPVLVPFPQSGIWTIAFVSGQVSNVKAALPKDGDYLSVYVPTTPNPTGGYY					
		130	140	150	160	170	180
		190	200	210	220	230	
60	orf98a.pep	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDL PVKTLAGPMPSEKADLPEQQX					
	orf98-1	IMVKKSDVRELDMSVDEALKYVISLGMVIPDDL PVKTLAGPMPSEKADLPEQQX					
		190	200	210	220	230	
65							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF98 shows 95.3% identity over a 233 aa overlap with a predicted ORF (ORF98ng) from *N.gonorrhoeae*:

		10	20	30	40	50	60	
5	orf98.pep	MTVTA	EEGGKAAKALKKYLI	TGILVWLPIA	VTWVVS	YIVSASDQ	LVNLLPKQWRPQYVL	60
	orf98ng	MTEPAA	EEGGKAAKALKKYLI	TGILVWLPIA	VTWVVS	YIVSASDQ	LVNLLPKQWRPQYVL	60
	orf98.pep	GFNIP	GLGVIAIAVLFVTGL	FAANVLGRQ	ILAAWDSLL	GRIPVVK	SIYSSVKKVSEYVL	120
10	orf98ng	GFNIP	GLGVIAIAVLFVTGL	FAANVLGRQ	ILAAWDSLL	XRIPVVK	SIYSSVKKVSESL	120
	orf98.pep	SDSSRS	FKTPVLVPPFPQGI	WTIAFVSGQV	SNVKAALP	XDGDYLS	SVYVPTTNPNTGGYY	180
	orf98ng	SDSSRS	FKTPVLVPPFPQGI	WTIAFVSGQV	SNVKAALP	QDGDYLS	SVYVPTTNPNTGGYY	180
15	orf98.pep	IMVKK	SDVRELDMSVDE	LKXVVISL	GMVIPDDL	PVKTLAX	PMPSEKADLPEQQ	233
	orf98ng	IMVKK	SDVRELDMSVDE	ALKYVISL	GMVIPDDL	PVKTLAG	PMPPEKAELPEQQ	233

20 The complete length ORF98ng nucleotide sequence <SEQ ID 745> is predicted to encode a protein having amino acid sequence <SEQ ID 746>:

	1	MTEPAAEGGK	AAKALKKYLI	TGILVWLPIA	VTWVVS	YIVSASDQ	LVNLL
	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFVTG	LFAANVLGRQ	ILAAWDSLLX	
	101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLVPPFQSGI	WTIAFVSGQV	
25	151	SNVKAALPQ	DGDYLSVYVP	TPNPTGGYY	IMVKKSDVRE	LDMSVDEALK	
	201	YVISLGMVIP	DDL	PVKTLAG	PMPPEKAELP	EQQ*	

Further work revealed the complete nucleotide sequence <SEQ ID 747>:

	1	ATGACGGAAC	CTGCGGCCGA	AGGCGGCAAA	GCTGCCAAGG	CGTTAAAAAA	
	51	ATATCTGATT	ACAGGCATTT	TGGTCTGGCT	GCCGATTGCG	GTAACGGTTT	
30	101	GGGTGGTTTC	CTATATCGTT	TCCGCGTCCG	ACCAGCTGTG	CAACCTGCTG	
	151	CCGAAGCAAT	GGCGGCCGCA	ATATGTTTTG	GGGTTTAATA	TCCCCGGGCT	
	201	CGGCGTTATT	GTTGCCATTG	CCGTATTGTT	TGTAACCGGA	TTATTTGCCG	
	251	CAAACGTGTT	GGCGCGGCAG	ATTCTTGCCG	CGTGGGACAG	CCTGTTg99g	
	301	c9gaTTCCGG	TGTCAAATC	CATCTATTCG	AGTGTGAAA	AAGTATCCGA	
35	351	ATCGCTGCTG	TCCGACAGCA	GCCGTTCTGT	TAAAACGCCG	GTA	CTCGTGC
	401	CGTTTCCCA	ATCGGTTATT	TGGACAATCG	CATTCGTGTC	CGGTGAGGTG	
	451	TGGAATGCGG	TTAAGGCCGC	ATTGCCGCGC	GATGGCGATT	ATCTTTCCGT	
	501	GTATGTCCCG	ACCACGCCCA	ACCCGACCGG	CGGTACTAT	ATTATGGTAA	
	551	AGAAAAGCGA	TGTGCGCGAA	CTCGATATGA	GCGTGGACGA	AGCGTTGAAA	
40	601	TATGTGATTT	CGCTGGGTAT	GGTCATCCCT	GACGACCTGC	CCGTCAAAAC	
	651	ATTGGCAGGA	CCTATGCCGC	CTGAAAAGGC	GGAGTTGCCC	GAACAACAAT	
	701	AA					

This corresponds to the amino acid sequence <SEQ ID 748; ORF98ng-1>:

	1	MTEPAAEGGK	AAKALKKYLI	TGILVWLPIA	VTWVVS	YIVSASDQ	LVNLL
45	51	PKQWRPQYVL	GFNIPGLGVI	VAIAVLFVTG	LFAANVLGRQ	ILAAWDSLLG	
	101	RIPVVKSIYS	SVKKVSESL	SDSSRSFKTP	VLVPPFQSGI	WTIAFVSGQV	
	151	SNVKAALPQ	DGDYLSVYVP	TPNPTGGYY	IMVKKSDVRE	LDMSVDEALK	
	201	YVISLGMVIP	DDL	PVKTLAG	PMPPEKAELP	EQQ*	

ORF98ng-1 and ORF98-1 show 97.9% identity in 233 aa overlap:

		10	20	30	40	50	60	
50	orf98-1.pep	MTEPAA	EEGGKAAKALKKYLI	TGILVWLPIA	VTWVVS	YIVSASDQ	LVNLLPKQWRPQYVL	
	orf98ng-1	MTEPAA	EEGGKAAKALKKYLI	TGILVWLPIA	VTWVVS	YIVSASDQ	LVNLLPKQWRPQYVL	
		10	20	30	40	50	60	
55	orf98-1.pep	GFNIP	GLGVIAIAVLFVTGL	FAANVLGRQ	ILAAWDSLL	GRIPVVK	SIYSSVKKVSESL	
		70	80	90	100	110	120	

-419-

orf98ng-1		GFNIPGLGVIVAI	AVLFTVGLFAAN	VLGRQILA	AWDSLLGRIP	PVVKSIYSSV	KKVSESL	
		70	80	90	100	110	120	
5	orf98-1.pep	SDSSRSFKTPVLV	PFPPQPGIWTIA	FVSGQVSN	AVKAALPKD	GDYLSVYV	PTTNPPTGG	YY
	orf98ng-1	SDSSRSFKTPVLV	PFPPQSGIWTIA	FVSGQVSN	AVKAALPKD	GDYLSVYV	PTTNPPTGG	YY
		130	140	150	160	170	180	
10	orf98-1.pep	IMVKKSDVRELD	MSVDEALKYVI	SLGMVIPDD	LPVKTLAG	PMPSEKAD	LPEQXX	
	orf98ng-1	IMVKKSDVRELD	MSVDEALKYVI	SLGMVIPDD	LPVKTLAG	PMPPEKAE	LPEQXX	
		190	200	210	220	230		

- 15 Based on this analysis, including the fact that the putative transmembrane domains in the gonococcal protein are identical to the sequences in the meningococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 89

- 20 The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 749>:

25	1	ATgAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CCGTCGGACT
	51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCCA	CGTGTATATC	GTA
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CTTGTTTAAA	TTCATTATCG	GgGgTACTCA
	201	ATATCCCCGA	AAAGATGCAG	CGTTTCGGTT	CGGCnCGTAA	AGGCCkCAAG
	251	ssCGsGCTTG	CCTTGAACAA	GGCGGGTTTG	GCGTATTTTG	AAGGGCGTTT
	301	TGAAAAGCG	GAAGTAGAAG	CCTCACGCGT	GTGGTCAAC	AAAGTAGGCC
	351	GgGAGACAAC	CGGACTTTGG	CATTGATGCT	GrGCGGCAC	GCCGCCGGAC
	401	AGATGGAAAA	CATCGAsTG	CGCGACCGTT	ATCTTGCGGA	AATCGCCAAA
	451	CTGCCGAAA	AACAGCAGCT	TTCCCGTTAT	CTTTTGTGG	CGGAATCGGC
	501	GTTGAACCG	CGCGATTACG	AAGCGGCGGA	AGCCAATCTT	CATGCGGCGG
	551	CGAAGATGAA	TGCCAACCTT	ACGCGCCTCG	TGCGTCTGCA	.ATTGCTTAC
	601	GCTTTTCGACA	GGGGCGACGC	GTTGCAGGTT	CTGGCAAAA	CCGAAAAACT
	651	TTCCAAGCG	GGCGGTTGG	GCAAATCGGA	AATGGAACGG	TATCAAAATT
	701	GGGCATATCC	GTCGCCAGCT	GGCGGATGCT	GCCGATGCCG	CCGCTTTGAA
	751	AACCTGCCTG	AAGCGGATTC	CCGACAGCCT	CAAAAACGGG	GAATTGAGCG
	801	TATCGGTTGC	GGAAAAGTAC	GAACGTTTGG	GACTGTATGC	CGATGCGGTC
	851	AAATGGGTCA	AACAGCATTA	TCCGCAsAAC	CGCCGCCCCG	AGCTTTTGGG
	901	AGCCTTTGTC	GAAAGCGTGC	GCTTTTGGG	CGAGCGCGAA	CAGCAGAAAG
	951	CCATCGATT	TGCCGATGCT	TGGCTGAAAG	AACAGCCCG	TAACGCCCTT
	1001	CTGCTGATGT	ATCTCGGTCG	GCTCGCCTTC	GGCCGCAAC	TTGGGGCAA
	1051	GGCAAAAGGC	TACCTTGAAG	CGAGCATTCG	ATTAAAGCCG	AGTATTTCCG
	1101	CGCGTTTGGT	TCTAACAAAG	GTTTTCGACG	AAATCGGAGA	ACCGCAGAAG
	1151	GCGGAGGCGC	AC...			

- 45 This corresponds to the amino acid sequence <SEQ ID 750; ORF100>:

50	1	MKTVVWIVVL	FAAAVGLALA	SGIYTGdVYI	VLGOTMLRIN	LHAFVLGSLI
	51	AVVVWYFLFK	FIIGVLNIPE	KMQRFSGSARK	GXXXXLALNK	AGLAYFEGRF
	101	EKAELASRV	LVNKVGRDNR	TLALMLXAHA	AGOMENIXXR	DRYLAEIAKL
	151	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	AAAKMNANLT	RLVRLXIRYA
	201	FDRGDALQVL	AKTEKLSKAG	ALGKSEMERY	QNWAYRRQLA	DAADAAALKT
	251	CLKRIPDSLK	NGELSVSVAE	KYERLGLYAD	AVKWVKQHYP	XNRRPELLEA
	301	FVESVRFLGE	REQQAIDFA	DAWLKEQPDN	ALLMYLGRL	AFGRKLWGKA
	351	KGYLEASIAL	KPSISARLVL	TKVFDEIGEP	QKAEAH...	

Further work revealed the complete nucleotide sequence <SEQ ID 751>:

55	1	ATGAAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CCGTCGGACT
	51	GGCGCTGGCT	TCGGGCATTT	ACACCGGCCA	CGTGTATATC	GTA
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CTTGTTTAAA	TTCATTATCG	GCGTACTCAA

201 TATCCCCGAA AAGATGCAGC GTTTCGGTTC GGCGCGTAAA GGCCGCAAGG
 251 CCGCGCTTGC CTTGAACAAG GCGGGTTTGG CGTATTTTGA AGGGCGTTTT
 301 GAAAAGGCGG AACTAGAAGC CTCACGCGTG TTGGTCAACA AAGAGGCCGG
 351 AGACAACCGG ACTTTGGCAT TGATGCTGGG CGCGCACGCC GCCGGACAGA
 5 401 TGGAAAACAT CGAGCTGCGC GACCGTTATC TTGCGGAAAT CGCCAAACTG
 451 CCGGAAAAAC AGCAGCTTTC CCGTTATCTT TTGTTGGCGG AATCGGCGTT
 501 GAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA
 551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCT
 601 TTCGACAGGG GCGACGCGTT GCAGGTTCTG GCAAAAACCG AAAAATTTC
 10 651 CAAGGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG
 701 CATACGCGCG CCAGCTGGCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC
 751 TGCCTGAAGC GGATTCCCGA CAGCCTCAA AACCAGGGAAT TGAGCGTATC
 801 GGTTCGGGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT GCGGTCAAAT
 851 GGGTCAAACA GCATTATCCG CACAACCGCC GCCCGAGCT TTTGGAAGCC
 15 901 TTTGTCGAAA GCGTGCCTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT
 951 CGATTTTGCC GATGCTTGGC TGAAAGAACA GCCCGATAAC GCGCTTCTGC
 1001 TGATGTATCT CGGTCGGCTC GCCTACGGCC GCAAACTTTG GGGCAAGGCA
 1051 AAAGGCTACC TTGAAGCGAG CATTGCATTA AAGCCGAGTA TTTCCGCGCG
 1101 TTTGGTTCTA GCAAAGGTTT TCGACGAAAT CGGAGAACC GAGAAGCGCG
 20 1151 AGCGCGCAGC CAACTTGTT TTGGAAGCCG TCTCCGATGA CGAACGTCAC
 1201 GCAGCGTTAG AGCAGCATAG CTGA

This corresponds to the amino acid sequence <SEQ ID 752; ORF100-1>:

1 MKTVVWIVVL FAAAVGLALA SGIYTGdVYI VLQOTMLRIN LHAFVLGSLI
 51 AVVVWYFLFK FIIGVNLNPE KMQRFGSARK GRKAALALNK AGLAYFEGRF
 25 101 EKAELEASRV LVNKEAGDNR TLALMLGAHA AGQMENIELR DRYLAEIAKL
 151 PEKQQLSRYL LLAESALNRR DYEAAEALNH AAKMNANLT RLVRQLRYA
 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQLA DAADAAALKT
 251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKQHYP HNRPELLEA
 301 FVESVRFLGE REQQAIDFA DAWLKEQPDN ALLMYLGR LAYGRKLWGKA
 30 351 KGYLEASIAL KPSISARLVL AKVFDEIGEP QKAEAQRLV LEAVSDDERH
 401 AALEQHS*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF100 shows 93.5% identity over a 386aa overlap with an ORF (ORF100a) from strain A of *N.*

35 *meningitidis*:

		10	20	30	40	50	60
orf100.pep		MKT	VV	WV	IV	VL	FAAAVGLALASGIYTGdVYI
orf100a		MKT	VV	WV	IV	VL	FAAAXGLALASGIYTGdVYI
		10	20	30	40	50	60
orf100.pep		FIIG	VNLN	PEK	MQR	FGSARK	GXKXXLALNKAGLAYFEGRF
orf100a		FIIG	VNLN	PEK	MQR	FGSARK	GRKAALALNKAGLAYFEGRF
		70	80	90	100	110	120
orf100.pep		TLAL	MLXA	HAAG	QMEN	IXRDR	YLAEIAKLPEKQQLSRYL
orf100a		TLAL	MLGA	HAAG	QMEN	IELRDR	YLAEIAKLPEKQQLSRYL
		130	140	150	160	170	180
orf100.pep		TLAL	MLXA	HAAG	QMEN	IXRDR	YLAEIAKLPEKQQLSRYL
orf100a		TLAL	MLGA	HAAG	QMEN	IELRDR	YLAEIAKLPEKQQLSRYL
		190	200	210	220	230	240
orf100.pep		AAAK	MNAN	LTRL	VLXIR	YAFDR	GDALQVLAKTEKLSKAG
orf100a		AAAK	MNAN	LTRL	VLXIR	YAFDR	GDALQVLAKTEKLSKAG
		250	260	270	280	290	300
orf100.pep		DAAD	AAALK	TCLK	RIPD	SLKNG	ELSVSVAEKYERLGLYAD
orf100a		DAAD	AAALK	TCLK	RIPD	SLKNG	ELSVSVAEKYERLGLYAD

-421-

		310	320	330	340	350	360
	orf100.pep	FVESVRFLGEREQK	ADFADAWLKEQPDN	ALLMYLGR	LAFGRKLWGK	KAKGYLEASIAL	
5	orf100a	FVESVRFLGERDQK	ADFADAWLKEQPDN	ALLXYLGR	LAYGRKLWGK	KAKGYLEASIAL	
		310	320	330	340	350	360
		370	380				
	orf100.pep	KPSISARLVLT	KVFEIGEPQKAEAH				
10	orf100a	KPSISARLVLA	KVFEIGEPQKAE	AQRNLVLAS	VAEENRPSAETHX		
		370	380	390	400		

The complete length ORF100a nucleotide sequence <SEQ ID 753> is:

15	1	ATGAAACGG	TAGTCTGGAT	TGTCGTCCTG	TTTGCCGCCG	CNNTCGGGCT
	51	GGCATTGGCG	TCGGGCATTN	ACACCGGCGA	CGTGTATATC	GTACTCGGAC
	101	AGACCATGCT	CAGAAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTCGCTGATT
	151	GCCGTCGTGG	TGTGGTATTT	CCTGTTCAAA	TTCATCATCG	CGCTACTCAA
	201	TANCCCCGAA	AAGATGCAGC	GTTTCGGTTC	GGCGCGTAA	GGCCGCAAGG
	251	CCGCGCTTGC	TTTGAACAAG	GCGGGTTTGG	CGTATTTTGA	AGGGCGTTTT
20	301	GAAAAGGCGG	AACCTGAAGC	CTCGCGCGTA	TTGGGAAACA	AAGAGGCGGG
	351	GGATAACCGG	ACTTTGGCAT	TGATGTTGGG	CGCACATGCC	GCCGGGCAGA
	401	TGGAACAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAAACTG
	451	CCGGAAAGC	AGCAGCTTTC	CCGTATCTT	TTGTTGGCGG	AATCGGCGTT
	501	GAACCGGCGC	GATTACGAAG	CGGCGGAAGC	CAATCTTCAT	GCGGCGGCCA
25	551	AGATGAATGC	CAACCTTACG	CGCCTCGTGC	GTCTGCAACT	TCGTTACGCT
	601	TTGCACAGGG	GCGACGCGTT	GCAGGTCTCTG	GCAAAACCGC	AAAAANTTTC
	651	CAAGGCGGGC	GCGTNGGGCA	AATCGGAAAT	GGAACGGTAT	CAAAATTGGG
	701	CATACCGCCG	CCAGCTGNCG	GATGCTGCCG	ATGCCGCCGC	TTTGAAACCC
	751	TGCCTGAAGC	GGATTCCCGA	CAGCCTCAAA	AACGGGGAAT	TGAGCGTATC
30	801	GGTTGCGGAA	AAGTACGAAC	GTTTGGGACT	GTATGCCGAT	GCGGTCAAAT
	851	GGGTCAAACA	GCATTATCCG	CACAACCGCC	GACCCGAACT	TTTGGAAGCN
	901	TTTGTGCGAA	GCGTGCGCTT	TTTGGGCGAA	CGCGATCAGC	AGAAAGCCAT
	951	CGATTTTGCC	GATGCTTGGC	TGAAAGAACA	GCCCCGATAAT	GCGCTTCTGC
	1001	TGANGTATCT	CGTCCGGCTC	GCCTACGGCC	GCAAACTTTG	GGGCAAGGCA
35	1051	AAAGGCTACC	TTGAAGCGAG	CATTGCATTA	AAGCCGAGTA	TTTCCGCGCG
	1101	TTTGGTTCTG	GCAAAGGTTT	TTGACGAAAC	CGGAGAACCG	CAGAAGGCGG
	1151	AGGCGCAGCG	CAACTTGGTT	TTGGCAAGCG	TTGCCGAGGA	AAACCGNCCT
	1201	TCCGCCGAAA	CCCATTGA			

This encodes a protein having amino acid sequence <SEQ ID 754>:

40	1	MKTVVWIVVL	FAAAXGLALA	SGIXTGDVYI	VLGQTMLRIN	LHAFVLGSLI
	51	AVVVWYFLFK	FIIGVLNXPE	KMQRFGSARK	GRKAALALNK	AGLAYFEGRF
	101	EKAELEASRV	LGNKEAGDNR	TLALMLGAHA	AGQMENIELR	DRYLAEIAKL
	151	PEKQQLSRYL	LLAESALNRR	DYEAEEANLH	AAAKMNANLT	RLVRLQLRYA
	201	FDRGDALQVL	AKTEKXSKAG	AXGKSEMERY	QNWAYRRQLX	DAADAAALKT
45	251	CLKRIPDSLK	NGELSVSAE	KYERLGLYAD	AVKWVKQHPY	HNRRPELLEA
	301	FVESVRFLGE	RDQKKAIDFA	DAWLKEQPDN	ALLXYLGR	AYGRKLWGKA
	351	KGYLEASIAL	KPSISARLV	LAKVFEDETGE	PQKAEQRNLV	LASVAEENRP
	401	SAETH*				

ORF100a and ORF100-1 show 95.1% identity in 406 aa overlap:

50		10	20	30	40	50	60
	orf100a.pep	MKTVVWIVVLFAAAXGLALAS	GIXTGDVYI	VLGQTMLRINLHAFVLGSLI	AVVVWYFLFK		
	orf100-1	MKTVVWIVVLFAAAGLALAS	GIYTGVDYI	VLGQTMLRINLHAFVLGSLI	AVVVWYFLFK		
55		10	20	30	40	50	60
	orf100a.pep	FIIGVLNXPEKMQRFGSARK	GRKAALALNKAGLAYFEGRF	EKAELEASRVLGNKEAGDNR			
	orf100-1	FIIGVLNIPKMQRFGSARK	GRKAALALNKAGLAYFEGRF	EKAELEASRVLVNKEAGDNR			
60		70	80	90	100	110	120
	orf100a.pep	TLALMLGAHAAGQMENIELR	DRYLAEIAKLPEKQQLSRYL	LLAESALNRRDYEAEEANLH			
	orf100-1	TLALMLGAHAAGQMENIELR	DRYLAEIAKLPEKQQLSRYL	LLAESALNRRDYEAEEANLH			
65		130	140	150	160	170	180
	orf100a.pep	TLALMLGAHAAGQMENIELR	DRYLAEIAKLPEKQQLSRYL	LLAESALNRRDYEAEEANLH			
	orf100-1	TLALMLGAHAAGQMENIELR	DRYLAEIAKLPEKQQLSRYL	LLAESALNRRDYEAEEANLH			

-422-

		130	140	150	160	170	180
5	orf100a.pep	190	200	210	220	230	240
	orf100-1	190	200	210	220	230	240
10	orf100a.pep	250	260	270	280	290	300
	orf100-1	250	260	270	280	290	300
15	orf100a.pep	310	320	330	340	350	360
	orf100-1	310	320	330	340	350	360
20	orf100a.pep	370	380	390	400		
	orf100-1	370	380	390	400		

Homology with a predicted ORF from *N.gonorrhoeae*

ORF100 shows 93.3% identity over a 386 aa overlap with a predicted ORF (ORF100ng) from *N.gonorrhoeae*:

30	orf100.pep	MKTVVWIVVLFAAAVGLALASGIYTG	60
	orf100ng	MKTVVWIVVLFAAAVGLALASGIYTG	60
35	orf100.pep	FIIGVLNIPEKMQRFGSARKGXKX	120
	orf100ng	FIIGVLNIPENMRRSGSARKGRKA	120
40	orf100.pep	TLALMLXAHAGOMENIXXRDRYLAE	180
	orf100ng	TLALMLGAHAGOMENIELRDRYLAE	180
45	orf100.pep	AAAKMNANLTRLVRLXIRYAFDRG	240
	orf100ng	AAAKMNANLTRLVRLQLRYAFDRG	240
50	orf100.pep	DAADAAALKTKLRIPDSLKNGELSV	300
	orf100ng	DAADAAALKTKLRIPDSLKNGELSV	300
55	orf100.pep	FVESVRFLGEREQQKAIDFADAWLKE	360
	orf100ng	FVESVRFLGEREQQKAIDFADSWLKE	360
	orf100.pep	KPSISARLVLTQVFDEIGEPQKAEAH	386
	orf100ng	KPSIPARLVLAQVFDETAQSQAQAE	405

The complete length ORF100ng nucleotide sequence <SEQ ID 755> is:

	1	ATGAAACGG	TAGTCTGGAT	TGTTGTCCTG	TTTGCCGCCG	CCGTCGGACT
60	51	GGCGCTGGCT	TCGGGCATT	ACACCGGCGA	CGTGATATC	GTACTCGGAC
	101	AGACCATGCT	CAGAATCAAC	CTGCACGCCT	TTGTGTTAGG	TTGCTGATT
	151	GCGCTCGTGG	TGTGGTATT	CCTGTTTAAA	TTCATCATCG	GCGTACTCAA
	201	TATCCCCGAA	AATATGCGGC	GTTCCGGTTC	GGCGCGGAAA	GGCCGCAAGG
	251	CCGCGCTTGC	CTTGAATAAG	GCGGGTTTGG	CGTATTTTGA	AGGGCGTTTT
	301	GAAAGGCGG	AACTCGAAGC	CTCTCGAGTG	TTGGGCAACA	AAGAGGCCGG
65	351	AGACAACCGG	ACTTTGGCAT	TGATGCTGGG	CGCGCACGCG	GCAGGACAGA
	401	TGGAAATAT	CGAGCTGCGC	GACCGTTATC	TTGCGGAAAT	CGCCAACTG

-423-

451 CCGGAAAAAC AGCAGCTTTC CCGCTATCTT CTGCTGGCGG AATCGGCGTT
 501 AAACCGGCGC GATTACGAAG CGGCGGAAGC CAATCTTCAT GCGGCGGCGA
 551 AGATGAATGC CAACCTTACG CGCCTCGTGC GTCTGCAACT TCGTTACGCC
 601 TTCGATCGGG GCGATGCGTT GCAGGTTCTG GCAAAAaccg AAAAATTTTC
 5 CAAGGCGGGC GCGTTGGGCA AATCGGAAAT GGAACGGTAT CAAAATTGGG
 701 CATACCGCCG CCAGATGGCG GATGCTGCCG ATGCCGCCGC TTTGAAAACC
 751 TGCCTGAAGC GGATTCCCGA CAGCCTCAAA AACGGGGAAT TGagcGTATC
 801 GGTTCGGGAA AAGTACGAAC GTTTGGGACT GTATGCCGAT GCGGTCAAAT
 851 GGGTCAAACA GCATTATCCG CACAACCGCC GCCCGGAGCT TTTGGAAGCC
 10 TTTGTCGAAA GCGTGCCTT TTTGGGCGAG CGCGAACAGC AGAAAGCCAT
 901 CGATTTTGCC GATTCTTGGC TGAAAGAACA GCCCGATAAC GCGTCTCTGC
 951 TGATGTATCT CGGCCGGCTC GCCTACGGCC GCAAACCTTG GGGTAAGGCA
 1001 AAAGGCTACC TTGAAGCGAG TATTGCACTG AAGCCGAGTA TTCCGGCGCG
 1051 TTTGGTGTG GCAAAGGTTT TTGACGAAAC CGCACAGTCG CAAAAGCCG
 1101 AAGCACAGCG CAACCTGGTT TTGGCAAGCG TTGCCGGGGA AAACCGCCCT
 1151 TCCGCCGAAA CCCGTTGA
 1201

This encodes a protein having amino acid sequence <SEQ ID 756>:

1 MKTVVWIVVL FFAAVGLALA SGIYTGVDYI VLQOTMLRIN LHAFVLGSLI
 20 51 AVVVWYFLFK FIIGVLNIPE NMRRSGSARK GRKAALALNK AGLAYFEGRF
 101 EKAELEASRV LGNKEAGDNR TLALMLGAHA AGOMENIELR DRYLAEIAKL
 151 PEKQQLSRYL LLAESALNRR DYEAEEANLH AAAMNANLT RLVRQLRYA
 201 FDRGDALQVL AKTEKLSKAG ALGKSEMERY QNWAYRRQMA DAADAAALKT
 251 CLKRIPDSLK NGELSVSVAE KYERLGLYAD AVKWKQHYH HNRRPELLEA
 301 FVESVRELGE REQQKAIDFA DSWLKEQPDN ALLMYLGR LAYGRKLWGKA
 25 KGYLEASIAL KPSIPARLVL AKVFEDETAQS QKAEQRNLV LASVAGENRP
 401 SAETR*

ORF100ng and ORF100-1 show 95.3% identity in 402 aa overlap:

		10	20	30	40	50	60
30	orf100-1.pep	MKTVVWIVVLFAAAVGLALASGIYTGVDYIVLQOTMLRINLHAFVLGSLIAVVVWYFLFK					
	orf100ng	MKTVVWIVVLFAAAVGLALASGIYTGVDYIVLQOTMLRINLHAFVLGSLIAVVVWYFLFK					
		10	20	30	40	50	60
		70	80	90	100	110	120
35	orf100-1.pep	FIIGVLNIPEKMQRFGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLNKEAGDNR					
	orf100ng	FIIGVLNIPENMRRSGSARKGRKAALALNKAGLAYFEGRFEKAELEASRVLNKEAGDNR					
		70	80	90	100	110	120
40	orf100-1.pep	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH					
	orf100ng	TLALMLGAHAAGQMENIELRDRYLAEIAKLPEKQQLSRYLLLAESALNRRDYEAEEANLH					
		130	140	150	160	170	180
45	orf100-1.pep	AAAMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQLA					
	orf100ng	AAAMNANLTRLVRLQLRYAFDRGDALQVLAKTEKLSKAGALGKSEMERYQNWAYRRQMA					
		190	200	210	220	230	240
50	orf100-1.pep	DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWKQHYPHNRRPELLEA					
	orf100ng	DAADAAALKTCLKRIPDSLKNGELSVSVAEKYERLGLYADAVKWKQHYPHNRRPELLEA					
		250	260	270	280	290	300
55	orf100-1.pep	FVESVRELGEREQKKAIDFADAWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL					
	orf100ng	FVESVRELGEREQKKAIDFADAWLKEQPDNALLMYLGRLAYGRKLWGKAKGYLEASIAL					
		310	320	330	340	350	360
60	orf100-1.pep	KPSISARLVLAKEVFEDETAQSQKAEQRNLVLEAVSDDERHAALEQHSX					
	orf100n	KPSIPARLVLAKEVFEDETAQSQKAEQRNLVLEAVSDDERHAALEQHSX					
		370	380	390	400		

370 380 390 400

Based on this analysis, including the presence of a putative leader sequence, a putative transmembrane domain, and a RGD motif, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 90

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 757>

```

10      1  ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCTGT
      51  GTTTGCAGGG CTGTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
     101  TTGATGTGCC GCGCGGCAAT CCCGAGTATG TCGCTCTGTC GGGCATGGCG
     151  GTGCGGCTGT ACCGTTTAT GTCGCCGTG GGTTCGGCG CGGTCGTGTT
     201  CCGCGCGGCG ATACCGTTG CCGCGGCTG GTGGGCAGC GGCTGGGTAC
     251  ACGTCAAAC GTGTTTGGG TTGATGCTCT TGGCTTACCA GTTGTATTGC
     301  GCGGTGCTGC TCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
     351  CTGGTACCGC GTGTTCAACG AAATCCCGT GCTGCTGATG GTTGCCGCGC
     401  TGTATSTGGT CGTGTTCAAA CCGTTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 758; ORF102>:

```

20      1  MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA
      51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWS GWVHVKLCLG LMLLAYQLYC
     101  GVLLRRFDY SNAFSHRWYR VFNEIPVLLM VAALYXVVEK PF*

```

Further work revealed the complete nucleotide sequence <SEQ ID 759>:

```

25      1  ATGATGTTTT CTTGGTTCAA GCTGTTTCAC TTGTTTTTTG TCATTTCTGT
      51  GTTTGCAGGG CTGTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
     101  TTGATGTGCC GCGCGGCAAT CCCGAGTATG TCGCTCTGTC GGGCATGGCG
     151  GTGCGGCTGT ACCGTTTAT GTCGCCGTG GGTTCGGCG CGGTCGTGTT
     201  CCGCGCGGCG ATACCGTTG CCGCGGCTG GTGGGCAGC GGCTGGGTAC
     251  ACGTCAAAC GTGTTTGGG TTGATGCTCT TGGCTTACCA GTTGTATTGC
     301  GCGGTGCTGC TCGCCGTTT TCAGGATTAC AGCAATGCTT TTTCACACCG
     351  CTGGTACCGC GTGTTCAACG AAATCCCGT GCTGCTGATG GTTGCCGCGC
     401  TGTATCTGGT CGTGTTCAAA CCGTTTTGA

```

This corresponds to the amino acid sequence <SEQ ID 760; ORF102-1>:

```

35      1  MMFSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA
      51  VRLYRFMSPL GFGAVVFGAA IPFAAGWWS GWVHVKLCLG LMLLAYQLYC
     101  GVLLRRFDY SNAFSHRWYR VFNEIPVLLM VAALYLVVEK PF*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with HP1484 hypothetical integral membrane protein of *H. pylori* (accession number AE000647)

ORF102 and HP1484 show 33% aa identity in 143aa overlap:

```

40      orf102  3  FSWFKLFH LFFVISWFAG LFYLPRI FVN MAMIDVPRGN PEYVRLSGMA VRLYRFMSPL GF 62
      HP1484  8  FLWVKAFH VIAVISWMAAL FYLPRL FVHAENAHKKE FVG VVQIQEK--KLYSFIASPA M 65
      orf102  63  GAVVFGAA IPFAAG---WWSGWVHVKLCLGLMLLAYQLYCGVLLRRFDYSNAFSHRWY 119
      HP1484  66  GFTLITGILMLLIEPTLFKSGGWLHAKLALVLLLAYHFYCKKCMRELEKOPTRRNARFY 125
      orf102  120 RVFNEIPXXXXXXXXXXXXXFKPF 142
      HP1484  126 RVFNEAPTILMILIVILVVVKPF 148

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF102 shows 99.3% identity over a 142aa overlap with an ORF (ORF102a) from strain A of *N. meningitidis*:

5		10	20	30	40	50	60
	orf102.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL			
	orf102a	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL			
10		10	20	30	40	50	60
	orf102.pep	GFGAVVFGAAIPFAAGWWSGVHVKLC	LGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
	orf102a	GFGAVVFGAAIPFAAGWWSGVHVKLC	LGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
15		70	80	90	100	110	120
	orf102.pep	GFGAVVFGAAIPFAAGWWSGVHVKLC	LGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
	orf102a	GFGAVVFGAAIPFAAGWWSGVHVKLC	LGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
20		130	140				
	orf102.pep	VFNEIPVLLMVAALYLVVFKPFX					
	orf102a	VFNEIPVLLMVAALYLVVFKPFX					

The complete length ORF102a nucleotide sequence <SEQ ID 761> is:

	1	ATGATGTTTT	CTTGGTTCAA	GCTGTTTAC	TTGTTTTTG	TCATTTCTGTG
	51	GTTCAGGG	CTGTTTACC	TGCCGAGGAT	TTTCGTCAAT	ATGGCGATGA
25	101	TTGATGTGCC	GCGCGGCAAT	CCCGAGTATG	TGCGTCTGTC	GGGCATGGCG
	151	GTGCGGCTGT	ACCGTTTTAT	GTCCCGCTTG	GGCTTCGGCG	CGGTCGTGTT
	201	CGGCGCGGCG	ATACCGTTTG	CCGCCGGCTG	GTGGGGCAGC	GGCTGGGTAC
	251	ACGTCAAAC	GTGTTGGGC	TTGATGCTCT	TGGCTTACCA	GTTGTATTGC
	301	GGCGTGCTGC	TGCGCCGTTT	TCAGGATTAC	AGCAATGCTT	TTTCACACCG
30	351	CTGGTACCGC	GTGTTCAACG	AAATCCCCGT	GCTGCTGATG	GTTGCCGCGC
	401	TGTATCTGGT	CGTGTTCAAA	CCGTTTTGA		

This encodes a protein having amino acid sequence <SEQ ID 762>:

	1	MMFSWFKLFH	LFFVISWFAG	LFYLPRI	FVNMAMIDVPRGN	PEYVRLSGMA
	51	VRLYRMSPL	GFGAVVFGAA	IPFAAGWWS	GVHVKLCIG	LMLLAYQLYC
35	101	GVLLRRFQDY	SNAFSHRWYR	VFNEIPVLLM	VAALYLVVFK	PF*

ORF102a and ORF102-1 show complete identity in 142 aa overlap:

		10	20	30	40	50	60
	orf102a.pep	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL			
40	orf102-1	MMFSWFKLFHLFFVISWFAGLFYLPRI	FVNMAMIDVPRGNPEYVRLSGMAVRLYR	FMSPL			
		10	20	30	40	50	60
	orf102a.pep	GFGAVVFGAAIPFAAGWWSGVHVKLC	LGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
45	orf102-1	GFGAVVFGAAIPFAAGWWSGVHVKLC	LGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
		70	80	90	100	110	120
	orf102a.pep	GFGAVVFGAAIPFAAGWWSGVHVKLC	LGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
	orf102-1	GFGAVVFGAAIPFAAGWWSGVHVKLC	LGLMLLAYQLYCGVLLRRFQDYSNAF	SHRWYR			
50		130	140				
	orf102a.pep	VFNEIPVLLMVAALYLVVFKPFX					
	orf102-1	VFNEIPVLLMVAALYLVVFKPFX					

55 Homology with a predicted ORF from *N.gonorrhoeae*

ORF102 shows 97.9% identity over a 142 aa overlap with a predicted ORF (ORF102ng) from *N. gonorrhoeae*:


```

    orf102.pep  MMFSWFKLFHLFFVISWFAGLFYLPRIFFVNMAMIDVPRGNPEYVRLSGMAVRLYRFMSPL  60
               |||
    orf102ng    MMFSWFKLFHLFFVISWFAGLFYLPRIFFVNMAMIDAPRGNPEYVRLSGMAVRLYRFMSPL  60
5
    orf102.pep  GFGAVVFGAAIPFAAGRWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR  120
               |||
    orf102ng    GFGAVVFGAAIPFAAGRWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR  120
10
    orf102.pep  VFNEIPVLLMVAALYXVVFKEPF  142
               |||
    orf102ng    VFNEIPVLLMVAALYLVVFKPF  142

```

The complete length ORF102ng nucleotide sequence <SEQ ID 763> is:

```

1  ATGATGTTTT CTGTTTCAA GCTGTTTTCAC TTGTTTTTTG TCATTTTCGTG
51 GTTTCAGGG CTGTTTACC TGCCGAGGAT TTTCGTCAAT ATGGCGATGA
15 101 TTGATGCGCC GCGCGGCAAT CCCGAGTATG TGCGCCTGTC GGGGATGGCG
    151 GTGCGGTTGT ACCGTTTTAT GTCGCCTTTG GGTTCGGCG CGGTCGTGTT
    201 CCGCGCGGCG ATACCGTTTG CCGCcggccg GTGGGCGagc ggctggGTTT
    251 ACGTCAAAC GTGTTTGGGC TTGATGCTCT TGGCTTATCA GTTGTATTGC
    301 GCGGTGCTGC TGCGCCGTT TCAGGATTAC AGCAATGCTT TTTCACACCG
20 351 CTGGTACCGC GTGTTCAAcg aATCCCCGT GCTGCTGATG GTTGCCGCGC
    401 TGTATCTGGT CGTGTTCAAA CCGTTTGA

```

This encodes a protein having amino acid sequence <SEQ ID 764>:

```

1  MMFSWFKLFH LFFVISWFAG LFYLPRIFFV NMAMIDAPRN PEYVRLSGMA
25 51 VRLYRFMSPL GFGAVVFGAA IPFAAGRWGS GWVHVKLCLG LMLLAYQLYC
    101 GVLLRRFQDY SNAFSHRWYR VFNEIPVLLM VAALYLVVFK PF*

```

ORF102ng and ORF102-1 show 98.6% identity in 142 aa overlap:

```

               10      20      30      40      50      60
    orf102-1.pep MMFSWFKLFHLFFVISWFAGLFYLPRIFFVNMAMIDVPRGNPEYVRLSGMAVRLYRFMSPL
30  orf102ng     MMFSWFKLFHLFFVISWFAGLFYLPRIFFVNMAMIDAPRGNPEYVRLSGMAVRLYRFMSPL
               10      20      30      40      50      60

               70      80      90      100     110     120
    orf102-1.pep GFGAVVFGAAIPFAAGRWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR
35  orf102ng     GFGAVVFGAAIPFAAGRWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFSHRWYR
               70      80      90      100     110     120

               130     140
    orf102-1.pep VFNEIPVLLMVAALYLVVFKPF
40  orf102ng     VFNEIPVLLMVAALYLVVFKPF
               130     140

```

In addition, ORF102ng shows significant homology to a membrane protein from *H. pylori*:

```

45  gi|2314656 (AE000647) conserved hypothetical integral membrane protein
    [Helicobacter pylori] Length = 148
    Score = 79.2 bits (192), Expect = 1e-14
    Identities = 50/147 (34%), Positives = 68/147 (46%), Gaps = 13/147 (8%)
50  Query: 3  FSWFKLFHLFFVISWFAGLFYLPRIFFVNMAMIDAPRGNPEYVRLSGMAVRLYRFMSPLGF 62
    F W K FH+ VISW A LFYLP+ FV A + V++ +LY F++
    Sbjct: 8  FLWVKAFHVIAVISWMAALFYLPRLFFVYHAENAHKKEFVGVVQIQEK--KLYSFIASPM 65
55  Query: 63  GAVVFGAAIP-----FAAGRWGSGWVHVKLCLGLMLLAYQLYCGVLLRRFQDYSNAFS 115
    G + + F +G GW+H KL L ++LLAY YC +R + +
    Sbjct: 66  GFTLITGILMLLIEPTLFKSG----GWLHAKLALVLLLAYHFYCKKCMRELEKDPTRRN 121
60  Query: 116 HRWYRVFNEIPXXXXXXXXXXXXFKPF 142
    R+YRVFNE P KPF
    Sbjct: 122 ARFYRVFNEAPTILMILIVILVVVKPF 148

```

Based on this analysis, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 91

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 765>:-

```

5      1 ATGGCAAAAA TGATGAAATG GGC GGCTGTT GCGGCGGTCG CGGCGGCAGC
51     GGT TTGGGGC GGATGGTCTT AACTGAAGCC CGAGCCGCAC GTGCTTGATA
101    TTACGGAAAC GGT CAGGCGC GGC // .....
//... ATTCGTTTA CGATTTTGTC CGAACCGGAT ACGCCGATTA AGGCGAAGCT
51     CGACAGCGTC GACCCCGGGC TGACCACGAT GTCGTGCGGC GGT TACAACA
10     GCAGTACGGA TACGGCTTCC AATGCGGTCT ACTATTATGC CCGTTCGTTT
151    GTGCCGAATC CGGACGGCAA ACTCGCCACG GGGATGACGA CGCAGAATAC
201    GGT TGAATC GACGGCGTGA AAAATGTGCT GATTATTCCG TCGCTGACCG
251    TGAAAAATCG CGGCGGCAAG GCGTTTGTGC GCGTGTGGG TGCGGACGGC
301    AAGGCGGCGG AACGCGAAAT CCGGACCGGT ATGAGAGACA GTATGAATAC
15     CGAAGTAAAA AGCGGGTTGA AAGAGGGGGA CAAAGTGCTC ATCTCCGAAA
401    TAACCGCCGC CGAGCAACAG GAAAGCGGCG AACGCGCCCT AGGCGGCCCC
451    CCGCGCCGAT AA

```

This corresponds to the amino acid sequence <SEQ ID 766; ORF85>:

```

20      1 MAKMMKWA AV AAVAAAAVWG GWS.LKPEPH VLDITETVRR G.....
51     .....
101    .....
151    .....
201    ..... I SFTILSEPDT
251    PIKAKLDSVD PGLTTMSSGG YNSSTDASN AVYYYARSFV PNPDKLATG
25     MTTQNTVEID GVKNVLIIPS LTVKNRGGKA FVRVLGADGK AAEREIRTGM
351    RDSMNTEVKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR*

```

Further work revealed the further partial nucleotide sequence <SEQ ID 767>:

```

30      1 ..GTATCGGTCG GCGCGCAGGC ATCGGGGCAG ATTAAGATAC TTTATGTCAA
51     ACTCGGGCAA CAGGTTAAAA AGGGCGATTG GATTGCGGAA ATCAATTGCA
101    CCTCGCAGAC CAATACGCTC AATACGGA AAATCCAAGTT GGAACCGTAT
151    CAGGCGAAGC TGGTGTGCGC ACAGATTGCA TTGGGCAGCG CGGAGAAGAA
201    ATATAAGCGT CAGGCGGCGT TATGGAAGGA AAACGCGACT TCCAAGAGG
251    ATTTGGAAG CGCGCAGGAT GCGTTTGCCG CCGCCAAGC CAATGTTGCC
301    GAGCTGAAGG CTTTAATCAG ACAGAGCAA ATTTCCATCA ATACCGCCGA
35     GTCGGAATTG GGCTACACGC GCATTACCGC AACGATGGAC GGCACGGTGG
401    TGGCGATTCT CGTGAAGAG GGGCAGACTG TGAACGCGGC GCAGTCTACG
451    CCGACGATTG TCCAATTGGC GAATCTGGAT ATGATGTTGA ACAAAATGCA
501    GATTGCCGAG GCGGATATTA CCAAGGTGAA GCGGGGCGAG GATATTTCTG
551    TTACGATTTT GTCCGAACCG GATACGCCGA TTAAGGCGAA GCTCGACAGC
40     GTCGACCCCG GGCTGACCAC GATGTCGTCG GCGGTTACA ACAGCAGTAC
651    GGATACGGCT TCCAATGCGG TCTACTATTA TGCCCGTTTCG TTTGTGCCGA
701    ATCCGACGCG CAACTCGCC ACGGGGATGA CGACGCAGAA TACGGTTGAA
751    ATCGACGGCG TGA AAAATGT GCTGATTAT CCGTCGCTGA CCGTGAAAAA
801    TCGCGGCGGC AAGGCGTTTG TGCGCGTGT GGGTGCGGAC GGCAAGGCGG
45     CGGAACGCGA AATCCGGACC GGTATGAGAG ACAGTATGAA TACCGAAGTA
901    AAAAGCGGGT TGAAGAGGGG GGACAAAGTG GTCATCTCCG AAATAACCGC
951    CGCCGAGCAA CAGGAAAGCG GCGAACGCGC CTAGGCGGC CCGCCGCGCC
1001   GATAA

```

This corresponds to the amino acid sequence <SEQ ID 768; ORF85-1>:

```

50      1 ..VSVGAQASGQ IKILYVKLGQ QVKKGDIAE INSTSQNTL NTEKSKLETY
51     QAKLVSAQIA LGSAEKKYKR QAALWKENAT SKEDLESAQD AFAAAKANVA
101    ELKALIRQSK ISINTAESEL GYTRITATMD GTVVAILVEE GQTVNAAQST
151    PTIVQLANLD MMLNKMQIAE GDITKVKAGQ DISFTILSEP DTFIKAKLDS
201    VDPGLTTMSS GGYNSSDTA SNAVYYARS FVPNPDKLA TGMTTQNTVE
55     IDGVKNVLI I PSLTVKNRGG KAFVRVLGAD GKAAEREIRT GMRDSMNTEV
301    KSGLKEGDKV VISEITAAEQ QESGERALGG PPRR*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF85 shows 87.8% identity over a 41aa overlap and 99.3% identity over a 153aa overlap with an ORF (ORF85a) from strain A of *N. meningitidis*:

5	orf85.pep	MAKMMKWA	10	20	30	40	
		MAKMMKWA	10	20	30	40	
	orf85a	MAKMMKWA	10	20	30	40	50 60
10	orf85.pep	80	90	100		
	orf85a	TIVQLANL	210	220	230	240	250 260
15	orf85.pep	110	120	130	140	150	160
	orf85a	GYNSSTD	270	280	290	300	310 320
20	orf85.pep	170	180	190	200	210	220
	orf85a	AFVRVLG	330	340	350	360	370 380
25	orf85a	AFVRVLG	330	340	350	360	370 380
30	orf85.pep	PRRX					
	orf85a	PRRX					

The complete length ORF85a nucleotide sequence <SEQ ID 769> is:

35	1	ATGGCAAAA	TGATGAAATG	GGCGGCTGTT	GCGGCGGTCG	CGGCGGCAGC
	51	GGTTTGGGG	GGATGGTCTT	ATCTGAAGCC	CGAGCCGCAG	GCTGCTTATA
	101	TTACGGAAAC	GGTCAGGCGC	GGCGACATCA	GCCGGACGGT	TTCTGCAACA
	151	GGGGAGATT	CGCCGTCCAA	CCTGGTATCG	GTCGGCGCGC	AGGCATCGGG
	201	GCAGATTAAG	AAACTTTATG	TCAAACCTCG	GCAACAGGTT	AAAAAGGCGC
	251	ATTGATGTC	GGAATCAAT	TCGACCTCGC	AGACCAATAC	GCTCAATACG
40	301	GAAAAATCCA	AATTGGAAAC	GTATCAGGCG	AAGCTGGTGT	CGGCACAGAT
	351	TGCATTGGGC	AGCGCGGAGA	AGAAATATAA	GCGTCAGGCG	GCGTTGTGGA
	401	AGGATGATGC	GACCGCTAAA	GAAGATTGTT	AAAGCGCACA	GGATGCGCTT
	451	GCCGCCGCCA	AAGCCAATGT	TGCCGAGCTG	AAGGCTCTAA	TCAGACAGAG
	501	CAAAATTTCC	ATCAATACCG	CCGAGTCGGA	ATTGGGCTAC	ACGCGCATT
45	551	CCGCAACGAT	GGACGGCACG	GTGGTGGCGA	TTCTCGTGGA	AGAGGGGCAG
	601	ACTGTGAACG	CGGCGCAGTC	TACGCCGACG	ATTGTCCAAT	TGGCGAATCT
	651	GGATATGATG	TTGAACAAA	TGCAGATTGC	CGAGGGCGAT	ATTACCAAGG
	701	TGAAGGCGGG	GCAGGATATT	TCGTTTACGA	TTTTGTCCGA	ACCGGATACG
	751	CCGATTAAGG	CGAAGCTCGA	CAGCGTCGAC	CCCGGGCTGA	CCACGATGTC
50	801	GTCGGGCGGC	TACACAGCA	GTACGGATAC	GGCTTCCAAT	GCGGTCTACT
	851	ATTATGCCCG	TTCGTTGTG	CCGAATCCGG	ACGGCAAACT	CGCCACGGGG
	901	ATGACGACGC	AGAATACGGT	TGAAATCGAC	GGTGTGAAAA	ATGTGCTGAT
	951	TATTCCGTCG	CTGACCGTGA	AAAATCGCGG	CGGCAGGGCG	TTTGTGCGCG
	1001	TGTTGGGTGC	AGACGGCAAG	GCGGCGGAAC	GCGAAATCCG	GACCGGTATG
55	1051	AGAGACAGTA	TGAATACCGA	AGTAAAAAGC	GGGTTGAAAG	AGGGGGACAA
	1101	AGTGGTCATC	TCCGAAATAA	CCGCCGCCGA	GCAGCAGGAA	AGCGCGCAAC
	1151	GCGCCCTAGG	CGGCCGCCCG	CGCCGATAA		

This encodes a protein having amino acid sequence <SEQ ID 770>:

60	1	MAKMMKWA	AAVAAAAVWG	GWSYLKPEPO	AAYITETVRR	GDISRTVSAT
	51	GEISPSNLVS	VGAQASGQIK	KLYVKLGQQV	KKGDILAEIN	STSQTNTLNT
	101	EKSKLETYQA	KLVSQAIALG	SAEKKYKRQA	ALWKDDATAK	EDLESQAQDAL
	151	AAAKANVAEL	KALIRQSKIS	INTAESELGY	TRITATMDGT	VVAILVEEGQ
	201	TVNAAQSTPT	IVQLANLDM	LNMQIAEGD	ITKVKAGQDI	SFTILSEPDT
	251	PIKAKLDSVD	PGLTTMSSGG	YNSSTDASN	AVYYIARSFV	PNPDGKLATG
65	301	MTTQNTVEID	GVKNVLIIPS	LTVKNRGGRA	FVRVLGADGK	AAEREIRTGM

351 RDSMNTTEVKS GLKEGDKVVI SEITAAEQQE SGERALGGPP RR*

ORF85a and ORF85-1 show 98.2% identity in 334 aa overlap:

5	orf85a.pep	30 40 50 60 70 80	PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQOVKKGDLIAE
	orf85-1		VSVGAQASGQIKKLYVKLGQOVKKGDLIAE
10	orf85a.pep	90 100 110 120 130 140	INSTSQNTNTLNTSKLETYQAKLVSAQIALGSAEKYKQQAALWKDDATAKEDLESAQD
	orf85-1		INSTSQNTNTLNTSKLETYQAKLVSAQIALGSAEKYKQQAALWKENATSKEDLESAQD
15	orf85a.pep	150 160 170 180 190 200	ALAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST
	orf85-1		AFAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST
20	orf85a.pep	210 220 230 240 250 260	PTIVQLANLDMMLNKMQIAEGDITKVKGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS
	orf85-1		PTIVQLANLDMMLNKMQIAEGDITKVKGQDISFTILSEPDTPIKAKLDSVDPGLTTMSS
25	orf85a.pep	270 280 290 300 310 320	GGYNSSTDASNAVYYYARSFVNPDPGKLATGMTTONTVEIDGVKNVLIIPSLTVKNRGG
	orf85-1		GGYNSSTDASNAVYYYARSFVNPDPGKLATGMTTONTVEIDGVKNVLIIPSLTVKNRGG
30	orf85a.pep	330 340 350 360 370 380	RAFVRVLGADGKAAEREIRTGMRDSMNTTEVKSGLKEGDKVVI SEITAAEQQESGERALGG
	orf85-1		KAFVRVLGADGKAAEREIRTGMRDSMNTTEVKSGLKEGDKVVI SEITAAEQQESGERALGG
35	orf85a.pep	390	PPRRX
	orf85-1		PPRRX

Figure 19D shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF85a..

Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF85 shows a high degree of identity with a predicted ORF (ORF85ng) from *N.gonorrhoeae*:

50	ORF85	1 MAKMMKWAAVA AAAAAVWGGWS.LKPEPHVLDITETVRRG..... 40
	ORF85ng	1 MAKMMKWAAVA AAAAAVWGGWSYKPEPQAAYITEAVRRGDISRTVSAT 50
55	ORF85ISFTILSEPDT 250
	ORF85ng	201 TVNAAQSTPTIVQLANLDMMLNKMQIAEGDITKVKGQDISFTILSEPDT 250
60	ORF85	251 PIKAKLDSVDPGLTTMSSGGYNSSTDASNAVYYYARSFVNPDPGKLATG 300
	ORF85ng	251 PIKAKLDSVDPGLTTMSSGGYNSSTDASNAVYYYARSFVNPDPGKLATG 300
65	ORF85	301 MTTONTVEIDGVKNVLIIPSLTVKNRGGKAFVRVLGADGKAAEREIRTGM 350
	ORF85ng	301 MTTONTVEIDGVKNVLIIPSLTVKNRGGKAFVRVLGADGKAEREIRTGM 350
	ORF85	152 RDSMNTTEVKSGLKEGDKVVI SEITAAEQQESGERALGGPPRR 393
	ORF85ng	351 KDSMNTTEVKSGLKEGDKVVI SEITAAEQQESGERALGGPPRR 393

The complete length ORF85ng nucleotide sequence <SEQ ID 771> is:

```

1 ATGGCAAAAA TGATGAAATG GCGGCTGTT GCGGCGGTCG CGGCGGCaac
51 GGTTCGGGGC GGATGGTCTT ATCTGAAGCC CGAACCAGCAG GCTGCTTATA
101 TTACGGAaac 9gTCAGGCGC GCGGATATCA GCCGGACGGT TTCGCGGACG
151 GgcgAGATTT CGCCGTCCAA CCTGGTATCG GTCGGCGCGC AGGCTTCGGG
201 GCAGATTAaa AAGCTTTATG TCAAACTCGG GCAACAGGTC AAAAAGGGCG
251 ATTTGATTGC GGAAATCAAT TCGACCACGC AGACCAACAC GATCGATATG
301 GAAAAATCCA AATTGGAAAC GTATCAGGCG AAGCTGGTGT CGGCACAGAT
351 TGCATTGGGC AGCGCGGAGA AGAAATATAA GCGTCAGGCG GCGTTGTGGA
401 AGGATGATGC GACCTCTAAA GAAGATTGG AAAGCGCGCA GGATGCGCTT
451 GCCGCCGCCA AAGCCAATGT TGCCGAGTTG AAGGCTTAA TCAGACAGAG
501 CAAAATTTC ATCAATACCG CCGAGTCGGA TTTGGGCTAC ACGCGCATTa
551 CCGCGACGAT GGACGGCACG GTGGTGGCGA TTCCCGTGGA AGAGGGGCGAG
601 ACTGTGAACG CGCGCGAGTC TACGCCGACG ATTGTCCAAT TGGCGAATCT
651 GGATATGATG TTGAACAAAA TGCAGATTGC CGAGGGCGAT ATTACCAAGG
701 TGAAGCGGG GCAGGATATT TCGTTTACGA TTTTGTCCGA ACCGGATACG
751 CCGATTAAGG CGAAGCTCGA CAGCGTCGAC CCCGGGCTGA CCACGATGTC
801 GTCGGGCGGC TACAACAGCA GTACGGATAC GGCTTCCAAT GCGGTCTATT
851 ATTATGCCCG TTGTTTTGTG CCGAATCCGG ACGGCAAACT CGCCACGGGG
901 ATGACGACGC AGAATACGGT TGAATCGAC GGTGTGAAAA ATGTGTTGCT
951 TATTCGTCG CTGACCGTGA AAAATCGCGG CGGCAAGGCG TTCGTACGCG
1001 TGTGGGTGTC GGACGGCAAG GCAGTGAAC GCGAAATCCG GACCGGTATG
1051 AAAGACAGTA TGAATACCGA AGTGAAAAGC GGGTTGAAAG AGGGGGACAA
1101 AGTGGTCATC TCCGAAATAA CCGCCGCCGA GCAGCAGGAA AGCGGCGAAC
1151 GCGCCCTAGG CGGCCCGCGC CGCCGATAA

```

This encodes a protein having amino acid sequence <SEQ ID 772>:

```

1 MAKMMKWAaV AAAAAAVWG GWSYLKPEPQ AAYITEAVRR GDISRTVSAT
51 GEISPSNLVS VQAQASGQIK KLYVKLGQV KKGDLIAEIN STTQNTIDM
101 EKSKLETYQA KLVSAQIALG SAEKKYKRQA ALWKDDATSK EDLESAQDAL
151 AAAKANVAEL KALIRQSKIS INTAESDLGY TRITATMDGT VVAIPVEEGQ
201 TVNAAQSTPT IVQLANLDM LNKMQIAEGD ITKVKAGQDI SFTILSEPDT
251 PIKAKLDSVD PGLTTMSSGG YNSSTDASN AVYYYARFV PNPDKLATG
301 MTTQNTVEID GVKNVLLIPS LTVKNRGGKA FVRVLGADGK AVEREIRTGM
351 KDSMNTVEKS GLKEGDKVVI SEITAEQQE SGERALGGPP RR*

```

ORF85ng and ORF85-1 show 96.1% identity in 334 aa overlap:

```

30 40 50 60 70 80
orf85ng PQAAYITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKLYVKLGQVKKGDLIAE
40 orf85-1 VSVGAQASGQIKLYVKLGQVKKGDLIAE
10 20 30

90 100 110 120 130 140
orf85ng INSTTQNTIDMEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKDDATSKEDLESAQD
45 orf85-1 INSTSQTNTLNTEKSKLETYQAKLVSAQIALGSAEKKYKRQAALWKENATSKEDLESAQD
40 50 60 70 80 90

150 160 170 180 190 200
orf85ng ALAAAKANVAELKALIRQSKISINTAESDLGYTRITATMDGTVVAIPVEEGQTVNAAQST
50 orf85-1 AFAAAKANVAELKALIRQSKISINTAESELGYTRITATMDGTVVAILVEEGQTVNAAQST
100 110 120 130 140 150

210 220 230 240 250 260
orf85ng PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPLTTMSS
55 orf85-1 PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPLTTMSS
160 170 180 190 200 210

270 280 290 300 310 320
orf85ng GGYNSSTDASNNAVYYYARFVFPNPDKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG
60 orf85-1 GGYNSSTDASNNAVYYYARFVFPNPDKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG
220 230 240 250 260 270

330 340 350 360 370 380

```

-431-

```

orf85ng      KAFVRVLGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVISETAAEQQESGERALGG
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
orf85-1      KAFVRVLGADGKAAEREIRTGMKDSMNTEVKSGLKEGDKVVISETAAEQQESGERALGG
                280      290      300      310      320      330

5
orf85ng      390
PPRRX
|||||
orf85-1      PPRRX

```

10 In addition, ORF85ng shows significant homology to an *E.coli* membrane fusion protein:

gi|1787104 (AE000189) o380; 27% identical (27 gaps) to 332 residues from membrane fusion protein precursor, MTRC_NEIGO SW: P43505 (412 aa) [Escherichia coli] Length = 380

15 Score = 193 bits (485), Expect = 2e-48
Identities = 120/345 (34%), Positives = 182/345 (51%), Gaps = 13/345 (3%)

```

Query: 29  PQAAITETVRRGDISRTVSATGEISPSNLVSVGAQASGQIKKLYVKLGQVKKGLIAE 88
          P  Y T  VR GD+ ++V ATG++          V VGAQ SGQ+K L V +G +VKK  L-
Sbjct: 41  PVPTYQTLIVRPGDLQQSVLATGKLDALRKVDVGAQVSGQLKTLVSAIGDKVKKDQLLG 100

```

```

Query: 89  INSTTQNTIDMEKSKLETYQAKLVSAQIALGSAEKYKQAAALWKDDATSKEXXXXXX 148
          I+   N L  ++ L  +A+  A+  L  A  Y RQ  L  +  A S++
Sbjct: 101 IDPEQAENQIKEVEATLMELRAQRQQAELKLARVTYSRQORLAQTKAVSQQDLDTAAT 160

```

```

Query: 149 XXXXXXXXXXXXXXXXIRQSKISINTAESDLGYTRITATMDGTVVVAIPVEEGQTVNAAQST 208
          I++++ S++TA+++L YTRI A M G V I  +GQTV AAQ
Sbjct: 161 EMAVQAQIGTIDAQIKRNQASLDTAKTNLDYTRIVAPMAGEVTQITTLOGQTVIAAQQA 220

```

```

Query: 209 PTIVQLANLDMMLNKMQIAEGDITKVKAGQDISFTILSEPDTPIKAKLDSVDPGLTMTSS 268
          P I+ LA++ ML K Q++E D+  +K GQ  FT+L +P T  + ++ V P
Sbjct: 221 PNILTLADMSAMLVKAQVSEADVIHLKPGQKAWFTVLGDPLTRYEGQIKDVL----- 273

```

```

Query: 269 GGYNSTDATASNAVYYARSFVNPDPGKLATGMTTQNTVEIDGVKNVLLIPSLTVKNRGG 328
          + +  ++A++YYAR  VNP+G L  MT Q  +++ VKNVL IP  + + G
Sbjct: 274 -----TPEKVNDAIFYARFEVNPENGLRLDMTAQVHIQLTDVKNVLTIPSLALGDPVG 328

```

```

Query: 329 KAFVRV-LGADGKAVEREIRTGMKDSMNTEVKSGLKEGDKVVISET 372
          +V L  +G+  ERE+  G ++  + E+  GL+ GD+VVI E
Sbjct: 329 DNRYKVKLLRNGETREREVTIGARNDTDVEIVKGLEAGDEVVIGE 373

```

40 Based on this analysis, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF85-1 (40.4kDa) was cloned in the pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 19A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 19B), FACS analysis (Figure 19C), and ELISA (positive result). These experiments confirm that ORF85-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 92

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 773>:

```

50      1  ..ATTCCCGCCA CGATGACATT TGAACGCAGC GGCAATGCTT ACAAATCGT
      51  TTCGACGATT AAAGTGCCGC TATACAATAT CCGTTTCGAG TCCGGCGGTA
     101  CGGTTGTCGG CAATACCCTG CACCCTACCT ACTATAGAGA CATACGCAGG
     151  GGCAAACTGT ATGCGGAAGc CAAATTCGCC GACgGcAGCG TAACTTACGG
     201  CAAAGCGGGC GAGAGCAAAA CCGAGCAAAG CCCCAAGGCT ATGGATTGT

```

5
251 TCACGCTTGC CTGGCACTTG GCGGCAAATG ACGCGAAACT CCCCCCGGGG
301 CTGAAATCA CCAACGGCAA AAACTTTAT TCCGTCGGCG GTTGAATAA
351 GGCGGGTACA GGAAAATACA GCATAGGCGG CGTGGAACC GAAGTCGTCA
401 AATATCGGGT GCGGCGCGGC GACGATGCGG TAATGTATT CTTCGCACCG
451 TCCCTGAACA ATATTCCGGC ACAAATCGGC TATACCGACG ACGGCAAAAC
501 CTATACGCTG AACTCAAAT CCGTGAGAT CAACGGCCAG GCAGCCAAAC
551 CGTAA

This corresponds to the amino acid sequence <SEQ ID 774; ORF120>:

10
1 ..IPATMTFERS GNAYKIVSTI KVPLYNIRFE SGGTVVGNTL HPTYYRDIRR
51 GKLYAEAKFA DGSVTYKAG ESKTEQSPKA MDLFTLAWQL AANDAKLPPG
101 LKITNGKKLY SVGGLNKAGT GKYSIGGVET EVVKYRVRRG DDAVMYFFAP
151 SLNNIPAQIG YTDGKTYTL KLKSVQINGQ AAKP*

Further work revealed the complete nucleotide sequence <SEQ ID 775>:

15
1 ATGATGAAGA CTTTTAAAAA TATATTTTCC GCCGCCATTT TGTCCGCCGC
51 CTGCGCGTGC GCGTATGCGG CAGGGCTGCC CCAATCCGCC GTGCTGCACT
101 ATTCCGGCAG CTACGGCATT CCCGCCACGA TGACATTTGA ACGCAGCGGC
151 AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
201 TTTTCAGTCC GGCGGTACGG TTGTCGGCAA TACCCTGCAC CCTACCTACT
251 ATAGAGACAT ACGCAGGGGC AAAGTGTATG CGGAAGCCAA ATTCGCCGAC
20
301 GGCAGCGTAA CTTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAAAGCCC
351 CAAGGCTATG GATTTGTTC CGCTTGCTG CGAGTTGGCG GCAAATGACG
401 CGAAACTCCC CCCGGGGCTG AAAATCACCA ACGGCAAAAA ACTTTATTCC
451 GTCGGCGGTT TGAATAAGGC GGTACAGGA AAATACAGCA TAGGCGGCGT
501 GGAAACCGAA GTCGTCAAT ATCGGTGCG GCGCGGCGAC GATGCGGTAA
25
551 TGTATTTCTT CGCACCCTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
601 ACCGACGACG GCAAAACCTA TACGCTGAAA CTCAAATCGG TGCAGATCAA
651 CGGCCAGGCA GCCAAACCGT AA

This corresponds to the amino acid sequence <SEQ ID 776; ORF120-1>:

30
1 MMKTFKNIFS AAILSAALPC AYAAGLPQSA VLHYSYSGYI PATMTFERSG
51 NAYKIVSTIK VPLYNIRFES GGTVVGNLH PTYYRDIRRG KLYAEAKFAD
101 GSVTYKGAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
151 VVGLNKAGTG KYSIGGVETE VVKYRVRRGD DAVMYFFAPS LNNIPAQIGY
201 TDDGKTYTLK LKSVQINGQA AKP*

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF120 shows 92.4% identity over a 184aa overlap with an ORF (ORF120a) from strain A of *N. meningitidis*:

				10	20	30
40	orf120.pep			IPATMTFERSGNAYKIVSTIKVPLYNIRFE		
	orf120a	SAAILSAAALPCAYAAGLPXSAVLHYSYSGYIPATXXXXXXNAXKIVSTIKVPLYNIRFE		:		
		10 20 30 40 50 60				
45	orf120.pep	40 50 60 70 80 90		SGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVTYKAGESKTEQSPKAMDFTLAWQL		
	orf120a	70 80 90 100 110 120		SGGTVVGNLHPTYYRDIRRGKLYAEAKFADGSVTYKAXXXXXXQSPKAMDFTLAWQL		
50	orf120.pep	100 110 120 130 140 150		AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE EVVKYRVRRGDDAVMYFFAP		
	orf120a	130 140 150 160 170 180		AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETE EVVKYRVRRGDDAVMYFFAP		
55						

```

              160      170      180
orf120.pep  SLNNIPAQIGYTDGKTYTLKLKSVQINGQAAKPX
              |||||
orf120a     SLNNIPAQIGYTDGKTYTLKLKSVQINGQAAKPX
5           190      200      210      220

```

The complete length ORF120a nucleotide sequence <SEQ ID 777> is:

```

1  ATGATGAAGA CTTTAAAAA TATATTTTCC GCGGCCATTT TGTCGCGCGC
51  CCTGCCGTGC GCGTATGCGG CAGGGCTGCC CNAATCCGCC GTGCTGCACT
101 ATTCCGCGCAG CTACGGCATT CCGGCCACNA NNANNTNNGN ACNNGNGNC
151 AATGCTTNCA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
201 TTTCGAGTCC GCGGTACGGG TTGTCGGCAA TACCCTGCAC CCTACCTACT
251 ATAGAGACAT ACGCAGGGGC AAATGTATG CGGAAGCCAA ATTCGCGGAC
301 GGCAGCGTAA CCTACGGCAA AGCGGNNNNN ANNNNNNNNG NGCAAAGCCC
351 CAAGGCTATG GATTTGTTCA CGCTTGCNTG GCAGTTGGCG GCAAATGACG
15 401 CGAAACTCCC CCGGGGCTG AAAATCACCA ACGGCAAAA ACTTTATTC
451 GTCGGCGGTT TGAATAAGGC GGTACAGGA AAATACAGCA TAGGCGCGT
501 GGAACCGAA GTCGTCAAAT ATCGGGTGGC GCGCGGCGAC GATGCGGTAA
551 TGTATTTCTT CGCACCCTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
601 ACCGACGACG GCAAACCTA TACGTGAAA CTCAAATCGG TGCAGATCAA
20 651 CGGCCAGGCA GCCAAACCGT AA

```

This encodes a protein having amino acid sequence <SEQ ID 778>:

```

1  MMKTFKNIFS AAILSALPC AYAAGLPXSA VLHYSYSGYI PATXXXXXXX
51  NAXKIVSTIK VPLYNIRFES GGTVVGNTLH PTYYRDIRRG KLYAEAKFAD
101 GSVTYGKAXX XXXXQSPKAM DLFTLAWQLA ANDAKLPPLG KITNGKKLYS
25 151 VGGLNKAGTG KYSIGGVETE VVKYRVRGD DAVMYFFAPS LNNIPAQIGY
201 TDDGKTYTLK LKSVQINGQA AKP*

```

ORF120a and ORF120-1 show 93.3% identity in 223 aa overlap:

```

              10      20      30      40      50      60
orf120a.pep  MMKTFKNIFSAAILSAALPCAYAAGLPXSAVLHYSYSGYIPATXXXXXXXNAXKIVSTIK
              |||||
orf120-1     MMKTFKNIFSAAILSAALPCAYAAGLPQSAVLHYSYSGYIPATMTFERSGNAYKIVSTIK
              10      20      30      40      50      60

              70      80      90      100     110     120
orf120a.pep  VPLYNIRFESGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYGKAXXXXXXQSPKAM
              |||||
orf120-1     VPLYNIRFESGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAM
              70      80      90      100     110     120

              130     140     150     160     170     180
orf120a.pep  DLFTLAWQLAANDAKLPPLGKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRGD
              |||||
orf120-1     DLFTLAWQLAANDAKLPPLGKITNGKKLYSVGGLNKAGTGKYSIGGVETE VVKYRVRGD
              130     140     150     160     170     180

              190     200     210     220
orf120a.pep  DAVMYFFAPSLNNIPAQIGYTDGKTYTLKLKSVQINGQAAKPX
              |||||
orf120-1     DAVMYFFAPSLNNIPAQIGYTDGKTYTLKLKSVQINGQAAKPX
              190     200     210     220

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF120 shows 97.8% identity over 184 aa overlap with a predicted ORF (ORF120ng) from *N.gonorrhoeae*:

```

55  orf120.pep  IPATMTFERSGNAYKIVSTIKVPLYNIRFE 30
              |||||
      orf120ng  SAAILSAALPCAYAARLPQSAVLHYSYSGYIPATMTFERSGNAYKIVSTIKVPLYNIRFE 69

      orf120.pep  SGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAMD LFTLAWQL 90
60  |||||
      orf120ng  SGGTVVGNTLHPAYYKDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAMD LFTLAWQL 129

```



```

orfl20.pep  AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETEVEVKYRVRRGDDAVMYFFAP 150
            |||||
orfl20ng    AANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETEVEVKYRVRRGDDTVTYFFAP 189
5
orfl20.pep  SLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKP 184
            |||||
orfl20ng    SLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKP 223

```

The complete length ORF120ng nucleotide sequence <SEQ ID 779> is:

```

10      1  ATGATGAAGA CTTTAAAAA TATATTTTCC GCCGCCATT TGTCCGCCGC
      51  CCTGCCGTGC GCGTATGCGG CAAGGCTACC CCAATCCGCC GTGCTGCACT
     101  ATTCCGGCAG CTACGGCATT CCCGCCACGA TGACATTGTA ACGCAGCGGC
     151  AATGCTTACA AAATCGTTTC GACGATTAAA GTGCCGCTAT ACAATATCCG
     201  TTTCGAATCC GCGGTACGG TTGTCGGCAA TACCCTGCAC CCTGCCTACT
     15  251  ATAAAGACAT ACGCAGGGGC AAATGTATG CGGAAGCCAA ATTCCGCCGAC
     301  GGCAGCGTAA CCTACGGCAA AGCGGGCGAG AGCAAAACCG AGCAAAGCCC
     351  CAAGGCTATG GATTTGTTC CGCTTGCTG GCAGTTGGCG GCAAATGACG
     401  CGAAACTCCC CCCGGGTCTG AAAATCACCA ACGGCAAAA ACTTTATTCC
     451  GTCGGCGGCC TGAATAAGGC GGTACGGGA AAATACAGCA TaggCGGCGT
     20  501  GGAAACCGAA GTCGTCAAAT ATCGGGTGC GCGCGGCGAC GATACGTTAA
     551  CGTATTCTT CGCACCGTCC CTGAACAATA TTCCGGCACA AATCGGCTAT
     601  ACCGACGACG GCAAAACCTA TACGCTGAAG CTCAAATCGG TGCAGATCAA
     651  CGGACAGGCC GCCAAACCGT AA

```

This encodes a protein having amino acid sequence <SEQ ID 780>:

```

25      1  MMKTFKNIFS AAILSALPC AYAARLPQSA VLHYSYSYGI PATMTFERSG
     51  NAYKIVSTIK VPLYNIRFES GGTVVGNTLH PAYYKDIRRG KLYAEAKFAD
     101  GSVTYGKAGE SKTEQSPKAM DLFTLAWQLA ANDAKLPPGL KITNGKKLYS
     151  VGGLNKAGTG KYSIGGVETE VKYRVRRGD DTVTYFFAPS LNNIPAIQIGY
     201  TDDGKTYTLK LKSVQINGQA AKP*

```

30 In comparison with ORF120-1, ORF120ng shows 97.8% identity in 223 aa overlap:

```

      10      20      30      40      50      60
orfl20-1.pep  MMKTFKNIFSAAILSAALPCAYAAGLPQSAVLHYSYSYGI PATMTFERSGNAYKIVSTIK
orfl20ng      MMKTFKNIFSAAILSAALPCAYAARLPQSAVLHYSYSYGI PATMTFERSGNAYKIVSTIK
35      10      20      30      40      50      60
      70      80      90      100     110     120
orfl20-1.pep  VPLYNIRFESGGTVVGNTLHPTYYRDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAM
orfl20ng      VPLYNIRFESGGTVVGNTLHPAYYKDIRRGKLYAEAKFADGSVTYGKAGESKTEQSPKAM
40      70      80      90      100     110     120
      130     140     150     160     170     180
orfl20-1.pep  DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETEVEVKYRVRRGD
orfl20ng      DLFTLAWQLAANDAKLPPGLKITNGKKLYSVGGLNKAGTGKYSIGGVETEVEVKYRVRRGD
45      130     140     150     160     170     180
      190     200     210     220
orfl20-1.pep  DAVMYFFAPSLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKPX
orfl20ng      DTVTYFFAPSLNNIPAIQIGYTDDGKTYTLKLKSVQINGQAAKPX
50      190     200     210     220

```

55 This analysis, including the presence of a putative leader sequence in the gonococcal protein suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 93

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 781>:

```

      1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC
    51 .GCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA
   101 CTCCGTTTGC GGTGCGGCG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC
   151 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGTCTGT
    5 201 GATGGTGTTC TCCTTGATTT TGTGTTGGC ATTATTGTTG ATTATCGTCC
   251 CTATGCTGGT CGGGCAGTTC AACAAATTGG CATCGCGCCT GCCCAATTA
   301 ATCGGTTTGA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG
   351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC
   401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG
  10 451 AGGCAGGGCG GCAATATT..

```

This corresponds to the amino acid sequence <SEQ ID 782; ORF121>:

```

      1 MYRRKGRGIK PWMGAGXAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV
    51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
   101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW LQHTGELSN ALKAWFPVLM
  15 151 RQGGNI..

```

Further work revealed the complete nucleotide sequence <SEQ ID 783>:

```

      1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG GTGCCGGTGC
    51 GCGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA
   101 CTCCGTTTGC GGTGCGGCG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC
   151 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGTCTGT
   201 GATGGTGTTC TCCTTGATTT TGTGTTGGC ATTATTGTTG ATTATCGTCC
   251 CTATGCTGGT CGGGCAGTTC AACAAATTGG CATCGCGCCT GCCCAATTA
   301 ATCGGTTTGA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG
   351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC
   401 ATACGGGAGA GTTGAGCAAC GCGCTTAAGG CGTGGTTTCC CGTTTTGATG
   451 AGGCAGGGCG GCAATATTGT CAGCAGTATC GGCAACCTGC TGCTGCTTCC
   501 CTTGCTGCTT TACTATTTC TGTGCGATTG GCAGCGGTGG TCGTGC GGCA
   551 TTGCCAAACT GGTTCGAGG CTTTTTGCCG GTGCTTATAC GCGCATTACA
   601 GGCAATTTGA ACGAGGTATT GGGCGAATTT TTGCGCGGGC AGCTTCTGGT
   651 AATGCTGATT ATGGGCTTGG TTTACGGTTT GGGATTGGTG CTGGTCGGGC
   701 TGGATTCCGG GTTGGCATC GGTATGCTTG CCGGTATTTT GGTGTTTGTG
   751 CCTTATCTCG GGGCGTTTAC GGGATTGCTG CTTGCCACCG TCGCCGCCTT
   801 GCTCCAGTTC GGTTCGTGGA ACGGCATCCT ATCGGTTTGG GCGGTTTGTG
   851 CCGTAGGACA GTTCTCGAA AGTTTTTCA TTACGCCGAA AATCGTGGGA
   901 GACCGTATCG GGCTGTCGCC GTTTTGGGTT ATCTTTTCGC TGATGGCGTT
  1001 CCGTAACCTT GGTCTTGCTT CCGGAGGGCG TGCAGAAATA TTTGCCGGCG
  1051 AGTTTTTACC GGGCAGGTA G

```

This corresponds to the amino acid sequence <SEQ ID 784; ORF121-1>:

```

   40 1 MYRRKGRGIK PWMGAGXAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV
      51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
   101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW LQHTGELSN ALKAWFPVLM
   151 RQGGNIVSSI GNLLLP LLL YYFLDQWRW SCGIKLVPR RFAGAYTRIT
   201 GNLEVLGEF LRQQLV LLI MGLVYGLGLV LVGLDSGFAI GMLAGILVVF
   45 251 PYLGAFTGLL LATVAALLOF GSWNGILSVW AVFAVGQFLE SFFITPKIVG
      301 DRIGLSPFWV IFSLMAFGQL MGFVGM LAGL PLAAVTLVLL REGVQKYFAG
      351 SFYRGR*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

50 ORF121 shows 98.7% identity over a 156aa overlap with an ORF (ORF121a) from strain A of *N. meningitidis*:

```

      10      20      30      40      50      60
or121.pep MYRRKGRGIKPWMGAGXAFAALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
      10      20      30      40      50      60
55 or121a MYRRKGRGIKPWMDAGAAFAALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
      10      20      30      40      50      60
      70      80      90     100     110     120
or121.pep ASASMSVMVFSLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV

```

-436-

```

|||||
orf121a  ASASMSVMVFSLILLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
          70      80      90      100     110     120

5
          130     140     150
orf121.pep EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNI
          |||||
orf121a    EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
          130     140     150     160     170     180

10
orf121a    SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLD SGFAI
          190     200     210     220     230     240

```

The complete length ORF121a nucleotide sequence <SEQ ID 785> is:

```

15 1 ATGTATCGGA GGAAAGGGCG GGGCATCAAG CCGTGGATGG ATGCCGGTGC
    51 GCGGTTTGCC GCCTTGGTCT GGCTGGTTTT CGCGCTCGGC GATACTTTGA
   101 CTCGGTTTGC GGTTCGGCGG GTGCTGGCGT ATGTATTGGA CCCTTTGGTC
   151 GAATGGTTGC AGAAAAAGGG TTTGAACCGT GCATCCGCTT CGATGTCTGT
   201 GATGGTGTTC TCCTTGATTT TGTGTTGGC ATTATTGTTG ATTATTGTCC
   251 CTATGCTGGT CGGGCAGTTC AACAATTTGG CATCGCGCCT GCCCCAATTA
   301 ATCGGTTTTA TGCAGAACAC GCTGCTGCCG TGGTTGAAAA ATACAATCGG
   351 CGGATATGTG GAAATCGATC AGGCATCTAT TATTGCGTGG CTTCAGGCGC
   401 ATACGGCGCA GTTGAGCAAC GCGCTAAGG CGTGGTTTCC GTTTTGATG
   451 AGGCAGGGCG GCAATATTGT CAGCAGTATC GGCAACCTGC TGCTGCTTCC
   501 CTTGCTGCTT TACTATTTCC TGCTGGATTG GCAGCGGTGG TCGTGCGGCA
   551 TTGCCAAACT GGTTCGAGG CGTTTTGCCG GTGCTTATAC GCGCATTACA
   601 GGCAATTTGA ACGAGGTATT GGGCGAATTT TTGCGCGGGC AGCTTCTGGT
   651 GATGCTGATT ATGGGTTTGG TTTACGGCTT GGGGTTGGTG CTGGTCGGGC
   701 TGGATTCGGG GTTTGCAATC GGTATGTTTG CCGGTATTTT GGTTTTTGTT
   751 CCCTATTTGG GCGCGTTTAC AGGACTGCTG CTGGCAACCG TCGCCGCCCT
   801 GCTCCAGTTC GGTTCGTGGA ACGGCATCTT GGCTGTTTGG GCGGTTTTTG
   851 CCGTAGGACA GTTCTCGAA AGTTTTTTCA TTACGCCGAA AATCGTGGGA
   901 GACCGTATCG GCCTGTCGCC GTTTTGGGTT ATCTTTTCGC TGATGGCGTT
   951 CGGGCAGCTG ATGGGCTTTG TCGGAATGTT GGCCGGATTG CTTTGGCCG
  1001 CCGTAACCTT GGTCTTGCTT CGCGAGGGCG TGCAGAAATA TTTGCCGGC
  1051 AGTTTTTACC GGGGCAGGTA G

```

This encodes a protein having amino acid sequence <SEQ ID 786>:

```

40 1 MYRRKGRGIK PWNDAGAAFA ALVWLVFALG DTLTPFAVAA VLAYVLDPLV
    51 EWLQKKGLNR ASASMSVMVF SLILLLALLL IIVPMLVGQF NNLASRLPQL
   101 IGFQMNTLLP WLKNTIGGYV EIDQASIIAW LQAHTGELSNA KAWFPVLM
   151 RQGGNIVSSI GNLLLLPLLL YYFLDWQRW SCGIAKLVPR RFAGAYTRIT
   201 GNLNEVLGEF LRGQLLVMLI MGLVYGLGLV LVGLDSGFAI GMVAGILVEF
   251 PYLGAFTGLL LATVAALLQF GSWNGILAVW AVFAVGQFLE SFFITPKIVG
   301 DRIGLSPFWV IFSLMAFGQL MGFVGMLAGL PLAAVTLVLL REGVQKYFAG
   351 SFYRGR*

```

45 ORF121a and ORF121-1 show 99.2% identity in 356 aa overlap:

```

          10      20      30      40      50      60
orf121a.pep MYRRKGRGIK PWNDAGAAFAALVWLVFALGDTLTPFAVA VLAYVLDPLVEWLQKKGLNR
          |||||
orf121-1     MYRRKGRGIK PWMGAGAAFAALVWLVFALGDTLTPFAVA VLAYVLDPLVEWLQKKGLNR
          10      20      30      40      50      60

          70      80      90      100     110     120
orf121a.pep ASASMSVMVFSLILLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
          |||||
orf121-1     ASASMSVMVFSLILLLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
          70      80      90      100     110     120

          130     140     150     160     170     180
orf121a.pep EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
          |||||
orf121-1     EIDQASIIAWLQAHTGELSNAKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
          130     140     150     160     170     180

          190     200     210     220     230     240
orf121a.pep SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVLVGLD SGFAI

```

ORF121 shows 97.4% identity over a 156 aa overlap with a predicted ORF (ORF121ng) from *N.gonorrhoeae*:

20	orf121.pep	MYRRKGRGIKPWMGAGXAFALVWLVFALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR	60
	orf121.ng	MYRRKGRGIKPWMGAGAAFAALVWLVLVALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR	60
25	orf121.pep	ASASMSVMVFSLLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV	120
	orf121.ng	ASASMSVMVFSLLILLALLLIIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV	120
30	orf121.pep	EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNI	156
	orf121.ng	EIDQASIIAWFQAHTGELSNALKAWFPVLMKQGGNIVSTIGNLLPPLLLYYFLLDWHRW	180

35

1	MYRRKGRGIK	PWMGAGAAFA	ALVWLVAALG	DTLTPFAVAA	VLAYVLDPLV
51	EWLQKQKGLNR	ASASMSVMVF	SLILLALL	IIVPMLVGQF	NNLASRLPQL
101	IGFMQNTLLP	WLKNTIGGYV	EIDQASTIAW	FQAHTEGELS	ALKAWPFVLM
151	KQGGNIVSTI	GNLLLPFLLL	YFLLDWHRW	SCGIPKLVPR	RFAGAYTRIT
201	GNLNKVGWKF	LRGQLLGETE	RGAVVCRVGR	ECWEGGGARS	RPSDDGWPRW
251	GGG*				

40	1	ATGTATCGGA	GAAAAGGACG	GGGCATCAAG	CCGTGGATGG	GTGCCGGCGC
	51	GGCGTTTGCC	GCCTTGGTCT	GGCTGGTTTA	CGCGCTCGGC	GATACTTTGA
	101	CTCCGTTTGC	GGTTGCGGCG	GTGCTGGCGT	ATGTGTTTGA	CCCTTTGGTC
	151	GAATGSTTGC	AGAAAAAGGG	TTTGAACCGT	GCATCCGCTT	CGATGCTCTGT
45	201	GATGGTGTTT	TCCTTGATTT	TGTTGTTGGC	ATATTGTTTG	ATTATTTGCC
	251	CTATGCTGTT	CGGGCAGTTC	AATAATTTGG	CATCTCGCCT	GCCCCAATTA
	301	ATCGGTTTTA	TGCAGAACAC	GCTGCTGCCG	TGTTTGAAAA	ATACAATCGG
	351	CGGATATGTG	GAATTCGATC	AGGCATCTAT	TATTGCGTGG	TTTCAGGCGC
50	401	ATACGGGCGA	GTTGAGCAAC	CGCGTTAAGG	CGTGGTTTCC	CGTTTGTATG
	451	AAACAGGGCG	GCAATATTGT	CAGCAGTATC	GGCAACCTGC	TGCTGCCGCC
	501	CTTGCTGCTT	TACTATTTC	TGCTGGATTG	GCACGCGTGG	TCGTCGCGCA
	551	TCGCCAAACT	GGTTCGAGG	CGTTTTGCCG	GTGCTTATAC	CGGCATTACG
55	601	GGTAATTTGA	ACGAGGTATT	GGGCGAATT	TTGCGCGGTC	AGCTCTGGT
	651	GATGCTGATT	ATGGGCTTGG	TTTACGGTTT	GGGATTGATG	CTAGTCGGAC
	701	TGGATTCCGG	ATTTGCCATC	GGTATGGTTG	CCGGTATTTT	GGTGTTTGTC
	751	CCCTATTTTG	GTGCGTTTAC	GGGATTGCTG	CTTGCCACTG	TTGCAGCCTT
60	801	GCTCCAGTTC	GTTTCTGGA	ACGGAATCTT	GGCTGTTTGG	GCGTGTTTTG
	851	CCGTCCGTCA	GTTTCTCGAA	AGTTTTTTCA	TTACGCCGAA	AATTGTAGGA
	901	GACCGTATCG	GCCTGTGCGC	GTTTTGGGTT	ATCTTTTCGC	TGATGGCGTT
	951	CGGAGAGCTG	ATGGGCTTGT	TCGGAATGTT	GGCCGGATTG	CCTTTGGCCG
60	1001	CCGTAACCTT	GGTCTTGCTT	CGCGAGGGCG	CGCAGAAATA	TTTTGCCCGC
	1051	AGTTTTTACC	GGGGCAGGTA	G		

This corresponds to the amino acid sequence <SEQ ID 790; ORF121ng-1>:

```

1 MYRRKGRGIK PWMGAGAAFA ALVWL VYALG DTLTPFAVAA VLAYVLDPLV
51 EWLQKKGLNR ASASMSVMVF SLILLALLL IIVPMLVGQF NNLASRLPQL
101 IGFMQNTLLP WLKNTIGGYV EIDQASIIAW FQAHTGELSN ALKAWFPVLM
5 KQGGNIVSSI GNLLPPLLL YYFLLDWQRW SCGIAKLVPR RFAGAYTRIT
201 GNLNEVLGEF LRQQLVMLI MGLVYGLGLM LVGLDSGFAT GMVAGILVFV
251 PYLGAFTGLL LATVAALLQF GSWNGILAVW AVFAVGQFLE SFFITPKIVG
301 DRIGLSPFWV IFSLMAFGEL MGFVGMLAGL PLAAVTLVLL REGAQKYFAG
351 SFYRGR*

```

10 ORF121ng-1 and ORF121-1 show 97.5% identity in 356 aa overlap:

```

15 orf121-1.pep      10      20      30      40      50      60
    MYRRKGRGIK PWMGAGAAFAALVWL VFA LGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR
    orf121ng-1      10      20      30      40      50      60
    MYRRKGRGIK PWMGAGAAFAALVWL VYALGDTLTPFAVAAVLAYVLDPLVEWLQKKGLNR

20 orf121-1.pep      70      80      90      100     110     120
    ASASMSVMVFSLILLALLL IIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV
    orf121ng-1      70      80      90      100     110     120
    ASASMSVMVFSLILLALLL IIVPMLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYV

25 orf121-1.pep      130     140     150     160     170     180
    EIDQASIIAWLQAHTGELSNALKAWFPVLMRQGGNIVSSIGNLLLLPLLLYYFLLDWQRW
    orf121ng-1      130     140     150     160     170     180
    EIDQASIIAWFQAHTGELSNALKAWFPVLMKQGGNIVSSIGNLLLLPLLLYYFLLDWQRW

30 orf121-1.pep      190     200     210     220     230     240
    SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLVGLVGLDSGFAT
    orf121ng-1      190     200     210     220     230     240
    SCGIAKLVPRRFAGAYTRITGNLNEVLGEFLRGQLLVMLIMGLVYGLGLMLVGLDSGFAT

35 orf121-1.pep      250     260     270     280     290     300
    GMLAGILVFVPYLGAFTGLLLATVAALLQFGSWNGILSVWAVFAVGQFLESFFITPKIVG
    orf121ng-1      250     260     270     280     290     300
    GMVAGILVFVPYLGAFTGLLLATVAALLQFGSWNGILAVWAVFAVGQFLESFFITPKIVG

40 orf121-1.pep      310     320     330     340     350
    DRIGLSPFWVIFSLMAFGQLMGFVGMLAGLPLAAVTLVLLREGVQKYFAGSFYRGRX
    orf121ng-1      310     320     330     340     350
    DRIGLSPFWVIFSLMAFGELMGFVGMLAGLPLAAVTLVLLREGAQKYFAGSFYRGRX

```

In addition, ORF121ng-1 shows homology to a permease from *H.influenzae*:

```

50 sp|P43969|PERM HAEIN PUTATIVE PERMEASE PERM HOMOLOG Length = 349
    Score = 69.9 bits (168), Expect = 2e-11
    Identities = 67/317 (21%), Positives = 120/317 (37%), Gaps = 7/317 (2%)

    Query: 26 VYALGDTLTPFAVAAVLAYVLDPLVEWL-QKKGLNRASASMSVMVFSXXXXXXXXXXXXV 84
    +Y GD + P +A VL+Y+L+ + +L Q R A++ + VP
    Sbjct: 32 IYFFGDLIAPLLIALVLSYLLLEIPINFLNQYLKCPMLATILIFGSFIFGLAAVFFLVLP 91

55 Query: 85 MLVGQFNNLASRLPQLIGFMQNTLLPWLKNTIGGYVE-IDQASIIAWFQAHTGELSNALK 143
    ML Q +L S LP + N WL N Y E ID + + + F + ++ +
    Sbjct: 92 MLWNQITISLLSDLPAMF----NKSNEWLLNLPKNYPELIDYSMVDSIFNSVREKILGFE 147
    Query: 144 AWFPVLMKQGGNIVSSIGNXXXXXXXXXXXXWQRWSCGIAKLVPRRFAGAYTRITGNL 203

60 Sbjct: 148 SAVKLSLASIMNLVSLGIYAFLVPLMMFFMLKDKSELLQGVSRFLPKNRNLAFXRWK-EM 206

    Query: 204 NEVLGEFLRGQXXXXXXXXXXXXXXXXXXXXSGFAIGMVAGILVFVPYXXXXXXXXXXXX 263
    + + ++ G+ + + + G+ V VPY
65 Sbjct: 207 QQQISNYINGKLEILIVTLITYIIFLIFGLNYPDLLAFVGLSVLVPIYIGAVIVTIPVA 266

```

Query: 264 XXXXXQFGSWNGILAVWAVFAVGQFLESFFITPKIVGDRIGLSFFWVIFSLMAFGELMGF 323
 QFG + FAV Q L+ + P + + + L P +I S++ FG L GF
 Sbjct: 267 LVALFQFGISPTFWYIIIAFAVSQLLDGNLLVPYLFSEAVNLHPLIIIIISVLIFGGLWGF 326

5 Query: 324 VGMLAGLPLAAVTLVLL 340
 G+ +PLA + ++
 Sbjct: 327 WGVFFAIPLATLVKAVI 343

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the two proteins, it is predicted that the proteins from *N.meningitidis* and
 10 *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 94

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 791>:

15 1 ..ACTGCTTTTT CGGCGGCGCT GCGCTTGAGT CCATCATGAC TCGTCATATT
 51 TTTGTCTTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
 101 TTTGCACGTC CTGCCC GCCG CGTTCAAATG CGTACCAGCA ATACCGCCGC
 151 CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG
 201 TTTTGCCTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG
 251 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTTGG GTTCTGTGC
 20 301 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC
 351 TTTGATACGC CGCACGCAAA AGGACGCGGC TGTCCGCATC TTTGAACTCT
 401 GCGGCGGTGT CCGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC
 451 GAGCAGCGCG TCGGTAACGG CGTGACGAG CGCATCGGCA TCGGAGTGTC
 501 CGAGCAGCCC TTTTCAAAT GGGATTTCAG CTCCGCCAAG TATCAG..

25 This corresponds to the amino acid sequence <SEQ ID 792; ORF122>:

1 ..TAFSAALRLS PSXLVIFLSF GKPYQQTAAI LTFFCTSCPP RSNAYQQYRR
 51 LRLYAFHPPE IAEFFVGFAF DVDARNVYAO IGGDVGTHLR NVRRECGFLC
 101 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQT CRT
 151 EQRVGNVQVQ RIGIGVSEQP FFKWDFNSAK YQ..

30 Further work revealed the complete nucleotide sequence <SEQ ID 793>:

1 ATATCGTACT GGGCAAGCAG TTCGCCGAT TTTTGGAAAG TAGATACCGC
 51 GCCTTTGATT TTTTGGCCGC TCTTACCCAA GGCTTCGATG AAAAAGTTGA
 101 TGGTCGAGCC GGTACCGATG CCGATATATT CATTTCGGG TACGAATTCG
 151 ACTGCTTTTT CGGCGGCGAT GCGCTTGAGT TCGTCTGTG TCGTCATATT
 201 TTTGTCTTTT GGGAAACCGT ATCAACAAAC AGCCGCCATC TTAACATTTT
 251 TTTGCACGTC CTGCCC GCCG CGTTCAAATG CGTACCAGCA ATACCGCCGC
 301 CTGCGCCTCT ATGCCTTCCA TCCGCCCGAG ATAGCCGAGT TTTTCGTTGG
 351 TTTTGCCTTT GATGTTGACG CACGAAATGT CTATGCCCAA ATCGGCGGCG
 401 ATGTTGGCAC GCATTTGCGG AATGTGCGGC GCGAGTTGG GTTCTGTGC
 451 AATCACGGTC GTATCGACAT TGACCGCCTG CCAACCCTGC GCCTGAACGC
 501 TTTGATACGC CGCACGCAAA AGGACGCGGC TGTCCGCATC TTTGAACTCT
 551 GCGGCGGTGT CCGGGAAATG GCTGCCGATA TCGCCCAAAC CTGCCGCACC
 601 GAGCAGCGCG TCGGTAACGG CGTGACGAG CGCATCGGCA TCGGAGTGTC
 651 CGAGCAGCCC TTTTCAAAT GGGATTTCAG CTCCGCCAAG TATCAGCTTT
 701 CTGCCTTCGG TCAGTTGGTG GACATCGTAG CCCTGTCCGA TACGGATGTT
 751 CGTCATCGTT TGTGTTCTG A

This corresponds to the amino acid sequence <SEQ ID 794; ORF122-1>:

1 ISYWASSSPD FLEVDAPLI FLPLLPKASM KKLMEVPVPM PIYSFSGTNS
 51 TAFSAAMRLS SSCVVIFLSF GKPYQQTAAI LTFFCTSCPP RSNAYQQYRR
 101 LRLYAFHPPE IAEFFVGFAF DVDARNVYAO IGGDVGTHLR NVRREFGFLC
 151 NHGRIDIDRL PTLRLNALIR RTQKDAAVRI FELCGGVGEM AADIAQT CRT
 201 EQRVGNVQVQ RIGIGVSEQP FFKWDFNSAK YQLSAFGQLV DIVALS DTDV
 251 RHRLCS*

Computer analysis of this amino acid sequence gave the following results:

ORF122 shows 94.0% identity over a 182aa overlap with an ORF (ORF122a) from strain A of *N. meningitidis*:

```

5      orf122.pep      10      20      30
                        TAFSAALRLSPSXLVIFLSFGKPYQQTAAI
                        |||||:| | :|||
orf122a      FLPLLPKASMKKLMVEPVMPMPYSFSGTNSTAFSAAMRLSSSCVVIFLSFGKPYQQTAAI
                        30      40      50      60      70      80

10     orf122.pep      40      50      60      70      80      90
                        LTFFCTSCPPERSNAYQQYRRLRLYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR
                        ||| ||||| ||||| ||||| ||||| |||||
orf122a      LTFFXTSCPPERSNPQQYRRLRLYAFHAPETIEFFVGFAFXVDARNVYAQIGGDVGTHLR
                        90      100     110     120     130     140

15     orf122.pep      100     110     120     130     140     150
                        NVRRECGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT
                        |:| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
orf122a      NMRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT
                        150     160     170     180     190     200

20     orf122.pep      160     170     180
                        EQRVGNGVQQRIGIGVSEQPFFKWDFNSAKYQ
                        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
25     orf122a      EQRVGNGVQQRIGIGVSEQPFFKWDFNSAKYQLSAFGQLVDIVALSDTDVRRRLCSX
                        210     220     230     240     250

```

The complete length ORF122a nucleotide sequence <SEQ ID 795> is:

	1	ATATCATATT	GGGCAAGCAG	TTCCTGGAT	TTTTTGAAG	TAGATACCGC
30	51	GCCTTTGATT	TTTTTGCCGC	TCTTACCCAA	GGCTTCGATG	AAAAAGTTGA
	101	TGCTCGAACC	GGTACCGATG	CCGATGTATT	CGTTTTCGGG	TACGAATTCG
	151	ACTGCNTTTT	CGGCGGCGAT	TCGCTTGTGT	TGCTTCTGTG	TCGTATATT
	201	TTTGTCTTTT	GGGAAACCGT	ATCAACAAAC	AGCCGCCATC	TTAACATTTT
	251	TTNNNACGTC	CTGCCCGCGC	CGTTCAAATC	CTTACCAGCA	ATACCGCCGC
35	301	CTGCGACTCT	ATGCCCTTCCA	TGCGCCCGAG	ATAACCGAGT	TTTTCGTTGG
	351	TTTTGCCTTT	GANGTTGACG	CACGAAATGT	CTATGCCCAA	ATCGGCGGCG
	401	ATGTTGGCAC	GCATTTGCGG	AATATGCGGC	GCGAGTTTGG	GTTTCTGTGC
	451	AATCACGGTG	GTATCGACAT	TGACCGCCTG	CCAACCTGCG	GCCTGAACGC
	501	TTTGATACGC	CGCACGCAAA	AGGACGCGGC	TGTCGCGATC	TTTGAACTCT
40	551	GCGCGCGTGT	CGGCGAAATG	GCTGCCGATA	TGCCCCAAAC	CTGCCGCACC
	601	GAGCAGCGCG	TCGGTAACGG	CGTGCAGCAG	CGCATCGGCA	TCGGAGTGTC
	651	CGAGCAGCCC	TTTTTCAAAT	GGGATTTCAA	CTCCGCCAAG	TATCAGCTTT
	701	CTGCCTTCGG	TCAGTTGGTG	GACATCGTAG	CCCTGTCGGA	TACGGATGTT
	751	CGTCATCGTT	TGTGTTCTCG	A		

This encodes a protein having amino acid sequence <SEQ ID 796>:

45	1	ISYWASSSLD	FLEVDTAPLI	FLPLLPKASM	KKLMVEPVPM	PMYSFSGTNS
	51	TAFSAAMRLS	SSCVVIFLSF	GKPYQQTAAI	LTFFXTSCPP	RSNPYQGYRR
	101	LRLYAEHAPE	ITFEFVGFAF	XVDARNVYAQ	IGGDVGTHLR	NMRREFQGLR
	151	NHGRIDIDRL	PTLRLNALIR	RTQKDAAVRI	FELCGGVGEM	AADIAQTCRT
	201	EQRVGNVQQ	RIGIGVSEQP	FFKWFDFNSAK	YQLSAFGQLV	DIVALSOTDV
50	251	RHRLCS*				

ORF122a and ORF122-1 show 96.9% identity in 256 aa overlap:

```

10      20      30      40      50      60
orf122a.pep  ISYWASSSLDFLEVDTAPLIFLPLLPKASMKKLMVEPVMPMYSFSGTNSTAFSAAMRLS
55          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
orf122-1     ISYWASSSPDFLEVDTAPLIFLPLLPKASMKKLMVEPVMPPIYSFSGTNSTAFSAAMRLS
          10      20      30      40      50      60

          70      80      90      100     110     120
orf122a.pep  SSCVVIFLSFGKPYQQTAAILTFEFTSCPPRSNPYQQYRRLRLYAFHAPETIEFFVGFAF
60          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
orf122-1     SSCVVIFLSFGKPYQQTAAILTFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAF

```

-441-

		70	80	90	100	110	120
		130	140	150	160	170	180
5	orf122a.pep	XVDARNVYAQIGGDVGTHLRNMRR	EFGLCNHGRIDIDRLPTLRLNALIR	TQKDAAVRI			
	orf122-1	DVDARNVYAQIGGDVGTHLRNVRR	EFGLCNHGRIDIDRLPTLRLNALIR	TQKDAAVRI			
		130	140	150	160	170	180
10	orf122a.pep	FELCGGVGEMAADIAQTCRTEQ	RVGNVQQRIGIGVSEQPF	FKWDFNSAKYQLSAFGQLV			
	orf122-1	FELCGGVGEMAADIAQTCRTEQ	RVGNVQQRIGIGVSEQPF	FKWDFNSAKYQLSAFGQLV			
		190	200	210	220	230	240
15	orf122a.pep	DIVALSDTDVHRRLCSX					
	orf122-1	DIVALSDTDVHRRLCSX					
		250					
20							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF122 shows 89.6% identity over a 182 aa overlap with a predicted ORF (ORF122ng) from *N.gonorrhoeae*:

25	orf122.pep	TAFSAALRLSPSXLVIFLSFGKPYQQTAAI	30
	orf122ng	FLPLLKASMKKLMVEPVMPMYSFSGTNSTAFSAAMRLSSSCVVIFLSFGKPYQQTAAI	80
	orf122.pep	LTFFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAFDVDARNVYAQIGGDVGTHLR	90
30	orf122ng	LTFFCTSWPPRSNPYQQYRRLRLYAFHPPEIAEFFVGFAFDIDARNIDTQIGGDVGTHLR	140
	orf122.pep	NVRRECGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRIFELCGGVGEMAADIAQTCRT	150
	orf122ng	NVRCEFGFLCNHGRIDIDHLPTLRLNALIRRTQKDAAVRIFELCGGVGKMAADVAQTCRT	200
35	orf122.pep	EQRVGNVQQRIGIGVSEQPF	FKWDFNSAKYQ 182
	orf122ng	EQRVGNVQQRVGIRMPEQPF	FKWDFNSAKYQLSAFGQLVDIVALSDDIRHRLCS 256

The complete length ORF122ng nucleotide sequence <SEQ ID 797> is:

40	1	ATGTCGTACC	GGGCAAGCAG	TCGCCGGAT	TTTTTGGAGG	TTGAAACCGC
	51	GCCTTTGATT	TTTTTACCGC	TTTTGCCCAA	GGCTTCGATG	AAGAAATTGA
	101	tgTTCGAACC	GgtaCCGATG	CCGATGTATT	CGTTTTTCGGG	TACGAATTCTG
	151	ACTGCTTTTT	CGGCGGCGAT	GCGCttgAgt	TCgtcttgcg	TcgTCATATT
	201	TTTAtccttt	gGGAaaccct	atcaAcaAAc	agccgccatC	TTAACATTTT
45	251	TTTGACGtc	ctggccgcgcg	cgttcaAATc	cgtaccaGca	ataccgcccgc
	301	ctgcgcctCT	AtgcCTTCCA	TCCGCCCGAG	ATAGCCGAGT	TTTTCGTTGG
	351	TTTTGCCTTT	GATattGACG	CACGAAATAT	CGatacCCa	atcggcgGCG
	401	ATGTTGGCAC	GCATTTGCGG	AATGTGCGGT	GCGAGTTTGG	GTTTCTGTGC
	451	AATCACGGTC	GTATCGACAT	TGACCACCTG	CCAACCCTGC	GCCTGAACGC
50	501	TTTGATACGC	CGCACGCAAA	AGGACGCGGC	TGTCCGCATC	TTTGAACCTCT
	551	GCGGCGGTGT	CGGAAAATG	GCTGCCGATG	TCGCCCAAAC	CTGCCGCACC
	601	GAGCAGCgcg	tcggttaaCGG	CGTGACGAG	cgcgTcgGCA	TCCGAATGCC
	651	CGAGCAGCCC	TTTTTCAAT	GGGATTTCAA	CTCCGCCAAG	TATCAGCTTT
	701	CTGCCTTCGG	TCAATTGGTG	GACATCGTAG	CCCTGTCCGA	TACGGATATT
55	751	CGTCATCGTT	TGTGTTCTCTG	A		

This encodes a protein having amino acid sequence <SEQ ID 798>:

	1	MSYRASSSPD	FLEVETAPLI	FLPLLKASM	KKLMVEPVPM	PMYSFSGTNS
	51	TAFSAAMRLS	SSCVVIFLSF	GKPYQQTAAI	LTFFCTSWPP	RSNPYQQYRR
60	101	LRLYAFHPPE	IAEFFVGFAF	DIDARNIDTQ	IGGDVGTHLR	NVRCEFGFLC
	151	NHGRIDIDHL	PTLRLNALIR	RTQKDAAVRI	FELCGGVGKM	AADVAQTCRT
	201	EQRVGNVQQ	RVGIRMPEQP	FFKWDFNSAK	YQLSAFGQLV	DIVALSDDI
	251	RHRLCS*				

ORF122ng and ORF122-1 show 92.6% identity in 256 aa overlap:

		10	20	30	40	50	60
	orf122-1.pep	ISYWASSSPDFLEVDTAPLIFLPLLPKASMKKLMVEPVPMPIYSFSGTNSTAFSAAMRLS					
5	orf122ng	MSYRASSSPDFLEVETAPLIFLPLLPKASMKKLMVEPVPMPIYSFSGTNSTAFSAAMRLS					
		10	20	30	40	50	60
		70	80	90	100	110	120
10	orf122-1.pep	SSCVVIFLSFGKPYQQTAAAILTFFCTSCPPRSNAYQQYRRLRLYAFHPPEIAEFFVGFAF					
	orf122ng	SSCVVIFLSFGKPYQQTAAAILTFFCTSWPPRSNPYQQYRRLRLYAFHPPEIAEFFVGFAF					
		70	80	90	100	110	120
		130	140	150	160	170	180
15	orf122-1.pep	DVDARNVYAQIGGDVGTHLRNVRREFGFLCNHGRIDIDRLPTLRLNALIRRTQKDAAVRI					
	orf122ng	DIDARNIDTQIGGDVGTHLRNVRCEFGFLCNHGRIDIDHLPPTLRLNALIRRTQKDAAVRI					
		130	140	150	160	170	180
		190	200	210	220	230	240
20	orf122-1.pep	FELCGGVGEMAADIAQTCTEQRVGNVQQRIGIGVSEQPFFKWDENSAKYQLSAFGQLV					
	orf122ng	FELCGGVGKMAADVAQTCTEQRVGNVQQRVGIRMPEQPFKWDENSAKYQLSAFGQLV					
		190	200	210	220	230	240
25		250					
	orf122-1.pep	DIVALSDDVRHRLCSX					
	orf122ng	DIVALSDDIRHRLCSX					
30		250					

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 95

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 799>:

```

35      1  ..GCCGGCGCGA GTGCGAACAA CATTTCGCGG CGTTTTCGGG AAACACCCGT
      51  CGCTGTCAGC GTTACCCTGA TCGGCACGGT ACTTGCCGTC ATGCTGCCCG
     101  TTACCGAATA TGAAACTTC CTGCTGCTTA TCGGCTCGGT ATTGCGCCG
     151  ATGGGGCGGA TTTTGATTGC CGACTTTTTC GTCTGAAAC GCGGTGA

```

This corresponds to the amino acid sequence <SEQ ID 800; ORF125>:

```

40      1  ..AGASANNISA RFAETPVAVS VTLLGTVLAV MLPVTEYENF LLLIGSVFAP
     51  MGGFDCRLFR LETA*

```

Further work revealed the complete nucleotide sequence <SEQ ID 801>:

```

      1  ATGTCGGGCA ATGCCTCTC TCCTTCATCT TCCTCCGCCA TCGGGCTGAT
     51  TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACGCTGC
    101  TTGCGCCTTT GGGCTGGCAG CGCGTCTGG CGGCTCTACT TTTGGGTCAT
    151  GCCGTCGGCG GCGCGCTGTT TTTGCGGGC GCGTATATCG GCGCACTGAC
    201  CCGACGCAGC TCGATGGAAA GCGTGCGCCT GTCGTTCCGC AAACGCGGTT
    251  CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG
    301  GTGATGATT ACGCCGCGC AACGGTCAGC TCCGCTTTGG GCAAAGTGTT
    351  GTGGGACGGC GAATCTTTT TCTGGTGGG ATTGGCAAAC GGCGCGCTGA
    401  TTGTGCTGTG GCTGGTTTC GCGCACGCA AAACAGGCGG GCTGAAACCC
    451  GTTTCGATGC TGCTGATGCT GTTGGCGGTT CTGTGGCTGA GTGCCGAAGT
    501  CTTTTCCACG GCAGGCAGCA CCGCCGACA GGTTCAGAC GGCATGAGTT
    551  TCGGAACGGC AGTCGAGCTG TCCGCCGTA TGCCGCTTTC CTGGCTGCCG
    601  CTTGCCGCCG ACTACACGCG CCACGCGCGC CGCCCGTTTG CGGCAACCCT
    651  GACGGCAACG CTCGCCTACA CGCTGACCGG CTGCTGGATG TATGCCTTGG
    701  GTTTGGCAGC GCGGTTGTT ACCGGAGAAA CCGACGTGGC AAAAATCCTG
    751  CTGGGCGCAG GTTTGGGTGC GGCAGGCATT TTGGCGGTGC TCCTCTCCAC

```

15

1	MSGNASSPSS	SSAIGLIWFG	AAVSIAEIST	GTLLAPLGWO	RGLAALLLGH
51	<u>AVGGALFFAA</u>	AYIGALTGRS	SMESVRLSFG	KRGSVLFVSA	NMLQLAGWTA
101	VMIYAGATVS	SALGKVLWDG	ESFVWWALAN	GALVILWLVF	GARKTGGGLKT
151	<u>VSMLLMLLAV</u>	LWLSAEVFST	AGSTAAQVSD	GMISFGTAVEL	SAVMPLSWLP
201	LAADYTRHAR	RFFAATITAT	LAYTLTGCWM	YALGLAAALF	TGETDVAKIL
251	<u>LGAGLGAAGI</u>	<u>LAVVLSTVTT</u>	TFLDAYSAGA	SANNISARFA	ETPVAVGVTL
301	<u>IGTVLAVMLP</u>	VTEYENFLLL	IGSVFAPMAA	VLIADFFVLK	RREEIEGFDF
351	<u>AGLVLWLAGF</u>	<u>ILYRFLSSG</u>	WESSIGLTAP	VMSAVAIATV	SVRLFFFKTQ
401	<u>SLQRNPS*</u>				

ORF125 shows 76.5% identity over a 51aa overlap with an ORF (ORF125a) from strain A of *N. meningitidis*:

35 The ORF125a partial nucleotide sequence <SEQ ID 803> is:

This encodes a protein having the partial amino acid sequence <SEQ ID 804>:

BNSDOCID: <WO 9924578A2.1>

-444-

101 VMIYAGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARKTGGLKT
 151 VSMLLMLLAV LWLSAEXFST AGSTAAXVXD GMSFGTAVEL SAVMPLSWLP
 201 LAADYTRHAR RPFATLTAT LAYTLTGCMW YALGLAAALF TGETDVAKIL
 251 LGAGLGAAGI LAVVLSTVTT TFLDAYSAGV SANNISAKLS EIPAVAVAV
 301 VGTLLAVLLP VTEYENFLLL IGSVFAPMAA VLIADFFVLK RREEIEG..

ORF125a and ORF125-1 show 94.5% identity in 347 aa overlap:

		10	20	30	40	50	60
10	orf125a.pep	MSGNASSXSSSA	IGLIWFGAAVSIAE	ISTGTLLAPLGW	QRLAALLGHAV	GALFFAA	
	orf125-1	MSGNASSPSSSA	IGLIWFGAAVSIAE	ISTGTLLAPLGW	QRLAALLGHAV	GALFFAA	
		10	20	30	40	50	60
15	orf125a.pep	70	80	90	100	110	120
	orf125-1	70	80	90	100	110	120
20	orf125a.pep	130	140	150	160	170	180
	orf125-1	130	140	150	160	170	180
25	orf125a.pep	190	200	210	220	230	240
	orf125-1	190	200	210	220	230	240
30	orf125a.pep	250	260	270	280	290	300
	orf125-1	250	260	270	280	290	300
35	orf125a.pep	310	320	330	340		
	orf125-1	310	320	330	340	350	360

Homology with a predicted ORF from *N.gonorrhoeae*

45 ORF125 shows 86.2% identity over a 65aa overlap with a predicted ORF (ORF125ng) from *N.gonorrhoeae*:

	orf125.pep		AGASANNISARFAETPVAVSVTLIGTVLAV	30
	orf125ng	KILLGAGLGITGILAVVLSTVTTTFLDTYSAGASANNISARFAEIPVAVGVTLIRTVLAV		308
50	orf125.pep	MLPVTEYENFLLLIGSVFAPM-GGFDCLFRLETA	64	
	orf125ng	MLPVTEYKNFLLIRSVFGPMAGGFDCLFCLKTA	343	

55 An ORF125ng nucleotide sequence <SEQ ID 805> was predicted to encode a protein having amino acid sequence <SEQ ID 806>:

	1	MSGNASSPSS	SAAIGLVWFG	AAVSIAEIST	GTLLAPLGWO	RGLAALLLGH
	51	AVGGALFFAA	AYIGALTGRS	SMESVRLSFG	KCGSVLFSVA	NMLQLAGWTA
101	101	VMIYVATVS	SALGKVLWDG	ESFVWWALAN	GALIVLWLVF	GARRTGGLKT
	151	VSMLLMLLAV	LWLSVEVFAS	SGTNAAPAVS	DGMTFGTAVE	LSAVMPLSWL
201	201	PLAADYTRQA	RRPFAATLTA	TLAYTLTGCMW	MYALGLAAL	FTGETDVAKI
	251	LLGAGLGITG	ILAVVLSTVT	TTFLDTYSAG	ASANNISARF	AEIPVAVGV
60	301	LIRTVLAVML	PVTEYKNFLL	LIRSVFGPMA	GGFDCLFCL	KTA*

Further work revealed the following gonococcal DNA sequence <SEQ ID 807>:

```

1  ATGTCGGGCA ATGCCTCCTC TCCTTCATCT TCCGCCGCCA TCGGGCTGGT
51 TTGGTTCGGC GCGGCGGTAT CGATTGCCGA AATCAGCACG GGTACGCTGC
101 TCGCCCCCTT GGGCTGGCAG CGCGGTCTGG CGGCCCTGCT TTTGGGTGAT
151 GCCGTCGGCG GCGCGCTGTT TTTTGGCGCG GCGTATATCG GCGCACTGAC
201 CGGACGCAGC TCATGGAAA GTGTGCGCCT GTCGTTCGGC AAATGCGGTT
251 CAGTGCTGTT TTCCGTGGCG AATATGCTGC AACTGGCCGG CTGGACGGCG
301 GTGATGATTT ACGTCGGCGC AACGGTCAGC TCCGCTTTGG GCAAAGTGTT
351 GTGGGACGGC GAATCCTTTG TCTGGTGGGC ATTGGCAAAC GCGCACTGA
401 TCGTGCTGTG GCTGGTTTTC GCGGCACGCA GAACGGGCGG GCTGAAAACC
451 GTTTCGATGC TGCTGATGCT GCTTGCCGTG TTGTGGTTGA GCGTCGAAGT
501 GTTCGCTTCG TCCGGCACA ACGCCGCGCC CGCCGTTTCA GACGGCATGA
551 CCTTCGGAAC GGCAGTCGAA CTGTCCGCGC TCATGCCGCT TTCCTGGCTG
601 CCGCTGGCCG CCGACTACAC GCGCCAAGCA CGCCGCCCGT TTGCGGCAAC
651 CCTGACGGCA ACGCTCGCCT ATACGCTGAC GGGCTGCTGG ATGTATGCTT
701 TGGGTTTGGC GGGGCTCTG TTTACCGGAG AAACCGACGT GGCGAAAATC
751 CTGTTGGCGC CGGGCTTGGG CATAACGGGC ATTCTGGCAG TCGTCCTCTC
801 CACCGTTACC ACAACGTTTC TCGATACCTA TTCCGCCGGC GCGAGTGCAG
851 ACAACATTTC CGCGCGTTT GCGGAAATAC CCGTCGCTGT CGGCGTTACC
901 CTGATCGGCA CGGTGCTTGC CGTCATGCTG CCCGTTACCG AATATAAAAA
951 CTTCCTGCTG CTTATCGGCT CGGTATTGTC GCGGATGGCG GCGGTTTTGA
1001 TTGCCGACTT TTTCGTCTTA AAACGGCGTG AGGAGATTGA AGGCTTTGAC
1051 TTGCCGGGAC TGGTTCTGTG GCTGGCAGGC TTCATCCTCT ACCGCTTCCT
1101 GCTCTCGTCC GGTGGGAAA GCAGCATCGG TCTGACCGCC CCCGTAATGT
1151 CTGCCGTTGC CATTGCCACC GTATCGGTAC GCCTTTTCTT TAAAAAACCC
1201 CAATCTTTAC AAAGGAACCC GTCATGA

```

This corresponds to the amino acid sequence <SEQ ID 808; ORF125ng-1>:

```

1  MSGNASSPSS SAAIGLVWFG AAVSIAEIST GTLLAPLGWQ RGLAALLLGH
51 AVGGALFFAA AYIGALTGRS SMESVRLSFG KCGSVLFSVA NMLQLAGWTA
101 VMIYVGATVS SALGKVLWDG ESFVWWALAN GALIVLWLVF GARRTGGLKT
151 VSMLLMLLAV LWLSVEVFAS SGTNAAPAVS DGMTFGTAVE LSAMVPLSWL
201 PLAADYTRQA RRPFAATLTA TLAYTLTGCV MYALGLAAL FTGETDVAKI
251 LLGAGLGITG ILAVVLSTVT TFLDYSAG ASANNISARF AEIPVAVGVT
301 LIGTVLAVML PVTEYKNELL LIGSVFAPMA AVLIADFFVL KRREEIEGFD
351 FAGLVLWLAG FILYRFLSS GWESSIGLTA PUMSAVAIAT VSVRLFFKKT
401 QSLQRNPS*

```

ORF125ng-1 and ORF125-1 show 95.1% identity in 408 aa overlap:

```

10      20      30      40      50      60
orfl25-1.pep  MSGNASSPSSSSAIGLIWFGAAVSIAEISTGTLAPLGWQRLAALLLGHAVGGALFFAA
40          |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
orfl25ng-1    MSGNASSPSSSSAIGLVWFGAAVSIAEISTGTLAPLGWQRLAALLLGHAVGGALFFAA
10      20      30      40      50      60

70      80      90      100     110     120
orfl25-1.pep  AYIGALTGRSSMESVRLSFGKRGSVLFSVANMLQLAGWTAVMIYAGATVSSALGKVLWDG
45          |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
orfl25ng-1    AYIGALTGRSSMESVRLSFGKCGSVLFSVANMLQLAGWTAVMIYVGATVSSALGKVLWDG
70      80      90      100     110     120

130     140     150     160     170     179
orfl25-1.pep  ESFVWWALANGALIVLWLVFGARKTGGLKTVSMLLMLLAVLWLSAEVFSTAGSTAAQ-VS
50          |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
orfl25ng-1    ESFVWWALANGALIVLWLVFGARRTGGLKTVSMLLMLLAVLWLSVEVFASSGTNAAPAVS
130     140     150     160     170     180

180     190     200     210     220     230     239
orfl25-1.pep  DGMSFGTAVELSAVMPLSWLPLAADYTRHARRPFAATLTATLAYTLTGCVMYALGLAAL
55          |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
orfl25ng-1    DGMTFGTAVELSAVMPLSWLPLAADYTRQARRPFAATLTATLAYTLTGCVMYALGLAAL
190     200     210     220     230     240

240     250     260     270     280     290     299
orfl25-1.pep  FTGETDVAKILLGAGLGAAGILAVVLSTVTTTFLDAYSAGASANNISARFAETPVAVGVT
65          |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
orfl25ng-1    FTGETDVAKILLGAGLGITGILAVVLSTVTTTFLDYSAGASANNISARFAEIPVAVGVT
250     260     270     280     290     300

```

		300	310	320	330	340	350	359
	orf125-1.pep	LIGTVLAVMLPVTEYENFLL	LIGSVFAPMAAVLIADFFVL	KRREEIEGDFAGLVWLAG				
5	orf125ng-1	LIGTVLAVMLPVTEYKNFLL	LIGSVFAPMAAVLIADFFVL	KRREEIEGDFAGLVWLAG				
		310	320	330	340	350	360	
		360	370	380	390	400		
10	orf125-1.pep	FILYRFLSSGWESSIGLTAPVMSA	VAIATVSVRLFFKKTQSLQRNPSX					
	orf125ng-1	FILYRFLSSGWESSIGLTAPVMSA	VAIATVSVRLFFKKTQSLQRNPSX					
		370	380	390	400			

Based on this analysis, including the presence of putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 96

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 809>:

	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
20	51	GTTGCAGCTT	GCAGAACAAG	GTTATCAGAT	TGCACTTTTC	GATAAAAGCT
	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TAGCCGCCGC	CATGCTCGCG
	151	CCTGCAGCGG	A.ACGGTCGA	AGCCACGCC	GAAGTGGTCA	GGCTGGGCAG
	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TATGGCACGG	GCAGGACAAG
25	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGT.ACAGGA
	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACCGC
	401	AACTCGGCGG	ACGTTTTTAA	GACGGCATCT	ACCTGCCGAC	CGAAGC.CAG
	451	CTCGACGGGC	GGCAATTATA	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GCCTGCAAG..

30 This corresponds to the amino acid sequence <SEQ ID 810; ORF126>:

	1	MTRIAILGGG	LSGRLTALQL	AEQGYQIALF	DKSCRRGEHA	AAYVAAAMLA
	51	PAAXTVEATP	EVVRLGRQSI	PLWRGIRCLR	NHTMMQENG	SLIVWHGQDK
	101	PLSSEFVRHL	KRGXTDDEI	VRWRADDIAE	REPQLGGRFX	DGIYLPTEXQ
	151	LDGRQLXSAL	ADALDELNVP	CHWEHECVPE	ACK...	

35 Further work revealed the complete nucleotide sequence <SEQ ID 811>:

	1	ATGACCCGTA	TCGCCATCCT	CGGCGGCGGC	CTCTCGGGAA	GGCTGACCGC
	51	GTTGCAGCTT	GCAGAACAAG	GTTATCAGAT	TGCACTTTTC	GATAAAAGCT
	101	GCCGCCGGGG	CGAACACGCC	GCCGCCTATG	TGCGCCCGC	CATGCTCGCG
	151	CCTGCAGCGG	AAGCGGTCGA	AGCCACGCC	GAAGTGGTCA	GGCTGGGCAG
40	201	GCAGAGCATC	CCGCTTTGGC	GCGGCATCCG	ATGCCGTCTG	AACACGCACA
	251	CGATGATGCA	GGAAAACGGC	AGCCTGATTG	TGTGGCACGG	GCAGGACAAG
	301	CCATTATCCA	GCGAGTTCGT	CCGCCATCTC	AAACGCGGCG	GCGTAGCGGA
	351	TGACGAAATC	GTCCGTTGGC	GCGCCGACGA	CATCGCCGAA	CGCGAACCGC
	401	AACTCGGCGG	ACGTTTTTCA	GACGGCATCT	ACCTGCCGAC	CGAAGGCCAG
45	451	CTCGACGGGC	GGCAAAATAT	GTCTGCACTT	GCCGACGCTT	TGGACGAACT
	501	GAACGTCCCC	TGCCATTGGG	AACACGAATG	CGTCCCCGAA	GGCTGCAAG
	551	CCCAATACGA	CTGGCTGATC	GACTGCCGCG	GCTACGGCGC	AAAAACCGCG
	601	TGGAACCAAT	CCCCGAGCA	CACCAGCACC	CTGCGCGGCA	TACGCGGCGA
	651	AGTGGCGCGG	GTTTACACAC	CCGAAATCAC	GCTCAACCGC	CCCGTGCGTC
50	701	TGCTCCATCC	GCGTTATCCG	CTCTACATCG	CCCCGAAAGA	AAACCCAGTC
	751	TTCGTATCG	GCGCGACCCA	AATCGAAAGC	GAAAGCCAAG	CCCCGCCAG
	801	CGTGCCTTCA	GGGTTGGAAC	TCTTGTCGCG	ACTCTATGCC	ATCCACCCCG
	851	CCTTCGGCGA	AGCCGACATC	CTCGAAATCG	CCACCGGCCT	GCGCCCGACG
	901	CTCAACCACC	ACAACCCCGA	AATCCGTTAC	AACCGCGCCC	GACGCCTGAT
55	951	TGAAATCAAC	GGCCTTTTCC	GCCACGGTTT	CATGATCTCC	CCCGCCGTAA
	1001	CCGCCGCGCG	CGCCAGATTG	GCAAGTGGC	TGTTTGACGG	AAAAGACGCG
	1051	CCCGAACGCG	ATAAAGAAAG	CGGTTTGGCG	TATATCCGAA	GACAAGATTA
	1101	A				

This corresponds to the amino acid sequence <SEQ ID 812; ORF126-1>:

```

      1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAMLA
    51  PAEAVEATP EVVRLGRQSI PLWRGIRCRL NTHMTMQENG SLIVWHGQDK
   101  PLSSEFVRHL KRGGVADDEI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
    5  151  LDGRQILSAL ADALDELNVP CHWEHECVPE GLQAQYDWLI DCRGYGAKTA
   201  WNQSPENTST LRGIERGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENHV
   251  FVIGATQIES ESQAPASVRS GLELLSALYA IHPAFGEADI LEIATGLRPT
   301  LNHHNPEIRY NRARRLIEIN GLFRHGFMIS PAVTAAARL AVALFDGDKDA
   351  PERDKESGLA YIRROD*

```

10 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF126 shows 90.0% identity over a 180aa overlap with an ORF (ORF126a) from strain A of *N. meningitidis*:

```

      10      20      30      40      50      60
15  orf126.pep  MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAYVAAAMLAPAAXTVEATP
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  orf126a      MTRIAILGGGLSGRLTALQLAEQGYQIALFDKSCRRGEHAAYVAAAMLAPAAEAVEATP
      10      20      30      40      50      60

      70      80      90     100     110     120
20  orf126.pep  EVVRLGRQSIPLWRGIRCRLNTHMTMQENGSLIVWHGQDKPLSSEFVRHLKRGGXTDDEI
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  orf126a      EVVRLGRQXIPLWRGIRCHLKT PAMMXENGSLIVWHGQDKPLSNEFVRHLKRGGVADDXI
      70      80      90     100     110     120

      130     140     150     160     170     180
25  orf126.pep  VRWRADDIAEREPLGGRFXDGIYLPTEXQLDGRQLXSALADALDELNVPCHWEHECVPE
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
  orf126a      VRWRADDIAEREPLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHWEHECAPE
      130     140     150     160     170     180
30

```

The complete length ORF126a nucleotide sequence <SEQ ID 813> is:

```

      1  ATGACCCGTA TCGCCATCCT CGGCGGCGGC CTCTCNGGAA GGCTGACCGC
    51  ACTGCAGCTT GCAGAACAAG GTTATCAGAT TGCACTTTTC GATAAAGGCT
   101  GCCGCGGGGG CGAACACGCC GCCGCCTATG TTGCCGCCGC CATGCTCGCG
   35  151  CCTGCGGCGG AAGCGGTCGA AGCCACGCCT GAAAGTGGTCA GGCTGGGCAG
   201  GCAGANCATC CCGCTTTGGC GCGGCATCCG ATGCCATCTG AAAACGCCTG
   251  CCATGATGCA NGAAAACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAA
   301  CCTTTATCCA ACGAGTTTGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA
   351  TGACNAAATC GTCCGTTGGC GCGCGACGA CATCGCCGAA CGCGAACCGC
   40  401  AACTCGGCGG ACGTTTTTCA GACGGCATCT ACCTGCCGAC CGAAGGCCAG
   451  CTCGACGGGG GGCAAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT
   501  GAACGTCCCC TGCCATTGGG AACACGAATG TGCCCCGAA GACTTGCAAG
   551  CCCAATACGA CTGGCTGATC GACTGCCGCG GCTACGGCGC AAAAACCGCG
   601  TGGAAACCAAT CCCCCGANNA NACCAGCACC CTGCGCGGCA TACGCGGCGA
   651  AGTGGCGCGG GTTTACACAC CCGAAATCAC GCTCAACCGC CCCGTGCGCC
   701  TGCTACACCC GCGCTATCCG CTNTACATCG CCCCAGAAAG AAACNCGTC
   751  TTCGTATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CACCTGCCAG
   801  CGTGCCTTCC GGGCTGGAAC TCTTATCCGC ACTCTATGCC GTCCACCCCG
   851  CTTTCGGCGA AGCCGACATC CTCGAAATCG CCACGGCCT GCGCCCCACG
   901  CTCAATCACC ACAACCCCGA AATCCGTTAC AACCGCGCCC GACGCTGAT
   951  TGAAATCAAC GGCCTTTTCC GCCACGGTTT CATGATCTCC CCCGCCGTAA
  1001  CCGCCGCCGC CGTCAGATTG GCACTGGCAC TGTGTTGACGG AAAAGANGCG
  1051  CCCGAACGCG ATGAAGAAAG CGGTTTGGCG TATATCCGAA GACAAGATTA
  1101  A

```

55 This encodes a protein having amino acid sequence <SEQ ID 814>:

```

      1  MTRIAILGGG LSGRLTALQL AEQGYQIALF DKGCRERGEHA AAYVAAAMLA
    51  PAEAVEATP EVVRLGRQXI PLWRGIRCHL KTPAMMXENG SLIVWHGQDK
   101  PLSNEFVRHL KRGGVADDXI VRWRADDIAE REPQLGGRFS DGIYLPTEGQ
   151  LDGRQILSAL ADALDELNVP CHWEHECAPE DLQAQYDWLI DCRGYGAKTA
   201  WNQSPXXTST LRGIERGEVAR VYTPEITLNR PVRLHPRYP LYIAPKENXV
   60

```

251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIATGLRPT
 301 LNHHNPEIRY NRARRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKXA
 351 PERDEESGLA YIRRQD*

ORF126a and ORF126-1 show 95.4% identity in 366 aa overlap:

5	orf126a.pep	10	20	30	40	50	60
	orf126-1	10	20	30	40	50	60
10	orf126a.pep	70	80	90	100	110	120
	orf126-1	70	80	90	100	110	120
15	orf126a.pep	130	140	150	160	170	180
	orf126-1	130	140	150	160	170	180
20	orf126a.pep	190	200	210	220	230	240
	orf126-1	190	200	210	220	230	240
25	orf126a.pep	250	260	270	280	290	300
	orf126-1	250	260	270	280	290	300
30	orf126a.pep	310	320	330	340	350	360
	orf126-1	310	320	330	340	350	360
35	orf126a.pep	YIRRQDX					
	orf126-1	YIRRQDX					
40							
45							

Homology with a predicted ORF from *N.gonorrhoeae*

ORF126 shows 90% identity over a 180 aa overlap with a predicted ORF (ORF126ng) from *N.gonorrhoeae*:

50	orf126.pep	60
	orf126ng	60
55	orf126.pep	120
	orf126ng	120
60	orf126.pep	180
	orf126ng	180

60 An ORF126ng nucleotide sequence <SEQ ID 815> was predicted to encode a protein having amino acid sequence <SEQ ID 816>:

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTROGEHA AAYVAAAMLA

51 PAEEAVEATP EVIRLGRQSI PLWRGIRCL NTLTMMQENG SLIVWHGQDK
 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ
 151 LDGRQILSAL ADALDELNVP CHWEHECAPO DLQAQYDWVI DCRGYGAKTA
 201 WNQSPHTST LRGIERGEVAR VYTPFITLNR PVRLHPRYP LYIAPKENHV
 251 SSSARPKSKA KAKPPPAYVP GWNSYPRSM STPPSAKPTS SKWRPGLRPT
 301 LNHHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA
 351 PERDEESGLA YIGRQD*

Further work revealed the following gonococcal DNA sequence <SEQ ID 817>:

1 ATGACCCGTA TCGCCGTCCT CGGAGGCGGC CTTTCCGGAA GGCTGACCGC
 51 ATTGCAGCTT GCAGAACAAAG GTTATCAGAT TGAACCTTTC GACAAGGGCA
 101 CCCGCCAAGG CGAACACGCC GCCGCTATG TTGCCGCCGC GATGCTCGCG
 151 CCGCGCGCGG AAGCGGTGCA GGCAACGCC GAAGTCATCA GGCTGGGCG
 201 GCAGAGCATT CCGCTTTGGC GCGGCATCCG ATGCCGTCTG AACACGCTCA
 251 CGATGATGCA GGAAAACGGC AGCCTGATTG TGTGGCACGG GCAGGACAAG
 301 CCATTATCCA GCGAGTTCGT CCGCCATCTC AAACGCGGCG GCGTAGCGGA
 351 TGACGAAATC GTCCGTTGGC GCGCCGATGA AATCGCCGAA CGCGAACCGC
 401 AACTCGGCGG ACGTTTTCAC GACGGCATCT ACCTGCCGAC CGAAGGCCAG
 451 CTCGACGGGC GGCAATATT GTCTGCACTT GCCGACGCTT TGGACGAACT
 501 GAACGTCCTT TGCCATTGGG AACACGAATG CGCCCCCAA GACCTGCAAG
 551 CCCAATACGA CTGGGTAATC GACTGCCGGG GCTACGCGGC GAAACCGCG
 601 TGGAAACCAAT CCCCCGAGCA CACAGCACC TTGCGCGGCA TACGCGCGCA
 651 AGTGGCGCGG GTTACACGC CCGAAATCAC GCTCAACCGC CCGTGCGCC
 701 TGCTGCACCC GCGCTATCCG CTCTACATCG CCCCAGAAAG AAACCACGTC
 751 TTCGTCATCG GCGCGACCCA AATCGAAAGC GAAAGCCAAG CCCCAGCCAG
 801 CGTACGTTCC GGGCTGGAAC TCTATCCGC GCTCTATGCC GTCCACCCCG
 851 CCTTCGGCGA AGCCGACATC CTCGAAATCG CCGCCGCGCT GCGCCCCACG
 901 CTAACACCAC ACAACCCCGA AATCCGCTAC AGCCGCGAAC GCCGCTCAT
 951 CGAAATCAAC GGCCTTTTCC GGCACGGCTT TATGATTCC CCGCCCGTAA
 1001 CCGCCGCCGC CGTCAGATTG GCAGTGGCAC TGTTTGACGG AAAAGACGCG
 1051 CCCGAACGTG ATGAAGAAAG CGTTTGGCG TATATCGAA GACAAGATTA
 1101 A

This corresponds to the amino acid sequence <SEQ ID 818; ORF126ng-1>:

1 MTRIAVLGGG LSGRLTALQL AEQGYQIELF DKGTROGEHA AAYVAAAMLA
 51 PAEEAVEATP EVIRLGRQSI PLWRGIRCL NTLTMMQENG SLIVWHGQDK
 101 PLSSEFVRHL KRGGVADDEI VRWRADEIAE REPQLGGRFS DGIYLPTEGQ
 151 LDGRQILSAL ADALDELNVP CHWEHECAPO DLQAQYDWVI DCRGYGAKTA
 201 WNQSPHTST LRGIERGEVAR VYTPFITLNR PVRLHPRYP LYIAPKENHV
 251 FVIGATQIES ESQAPASVRS GLELLSALYA VHPAFGEADI LEIAAGLRPT
 301 LNHHNPEIRY SRERRLIEIN GLFRHGFMS PAVTAAAVRL AVALFDGKDA
 351 PERDEESGLA YIGRQD*

ORF126ng-1 and ORF126-1 show 95.1% identity in 366 aa overlap:

10 20 30 40 50 60
 orf126-1.pep MTRIAVLGGGLSGRLTALQLAEQGYQIALFDKGCRRGEHAAYVAAAMLAPAAEAVEATP
 45 orf126ng-1 MTRIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHAAYVAAAMLAPAAEAVEATP
 10 20 30 40 50 60
 70 80 90 100 110 120
 orf126-1.pep EVVRLGRQSIPLWRGIRCLNTHHTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI
 50 orf126ng-1 EVIRLGRQSIPLWRGIRCLNTLTMMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEI
 70 80 90 100 110 120
 130 140 150 160 170 180
 orf126-1.pep VRWRADDIAEREPQLGGRFSDGIYLPTEGQDLGRQILSALADALDELNVPCHWEHECVPE
 55 orf126ng-1 VRWRADEIAEREPQLGGRFSDGIYLPTEGQDLGRQILSALADALDELNVPCHWEHECAPO
 130 140 150 160 170 180
 190 200 210 220 230 240
 orf126-1.pep GLQAQYDWLIDCRGYGAKTAWNQSPEHTSTLRGIERGEVARVYTPFITLNRPVRLHPRYP
 60 orf126ng-1 DLQAQYDWVIDCRGYGAKTAWNQSPEHTSTLRGIERGEVARVYTPFITLNRPVRLHPRYP
 190 200 210 220 230 240

-450-

		250	260	270	280	290	300
	orf126-1.pep	LYIAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAIHPAFGEADILEIATGLRPT					
5	orf126ng-1	LYIAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAVHPAFGEADILEIAGLRPT					
		250	260	270	280	290	300
	orf126-1.pep	LNHHNPEIRYNRRRLIEINGLFRHGFMI	SPAVTAAARLAVALFDGKDAPERDKESGLA				
10	orf126ng-1	LNHHNPEIRYSRERRLIEINGLFRHGFMI	SPAVTAAAVRLAVALFDGKDAPERDEESGLA				
		310	320	330	340	350	360
15	orf126-1.pep	YIRRODX					
	orf126ng-1	YIGRODX					

Furthermore, ORF126ng-1 shows homology to a putative *Rhizobium* oxidase flavoprotein:

	gi 2627327 (AF004408) putative amino acid oxidase flavoprotein [Rhizobium etli]
20	Length = 327
	Score = 169 bits (423), Expect = 3e-41
	Identities = 112/329 (34%), Positives = 163/329 (49%), Gaps = 25/329 (7%)
	Query: 3 RIAVLGGGLSGRLTALQLAEQGYQIELFDKGTROGEHXXXXXXXXXXXXXXXXXXXXX 62
25	Sbjct: 2 RILVNGAGVAGLTVAWQLYRHGFRVTLAERAGTVGA-GASGFAGGMLAPWCERESAEEPV 60
	Query: 63 IRLGRQSIPLWRGIRCLNLTMTMQENGSLIVWHGQDKPLSSEFVRHLKRGGVADDEIVR 122
30	Sbjct: 61 LTLGRLAADWWEAA-----LPGHVHRRGTLVVAGGRDTGELDRFSRRTS-GWEWLDEVA- 113
	Query: 123 WRADEIAEREPOLGGRFSDGIYLPTEGQLDGRQILSALADALDELNVPCHEHECAPQDL 182
35	Sbjct: 114 -----IAALEPDLAGRFRRALFFRQEAHLDPQALAAALAGLEDARMRLTLG---VVGES 165
	Query: 183 QAQYDWVIDCRGYGAKTAWNQSPHTSTLRGIRGEVARVYTPEITLNRPVRLHLHPRYPLY 242
40	Sbjct: 166 DVDHDRVVDCTGAA-----QIGRLPGLRGVRGEMLCVETTEVSLSRPVRLHLHPRHPIY 218
	Query: 243 IAPKENHVFVIGATQIESESQAPASVRSGLELLSALYAVHPAFGEADILEIAGLRPTLN 302
45	Sbjct: 219 IVPRDKNRFMVGATMIESDDGGPITARSLMELLNAAAYAMHPAFGEARVETGTAGVREAYP 278
	Query: 303 HHNPEIRYSRERRLIEINGLFRHGFMI 331
	Sbjct: 279 DNLP--RVTQEGRTLHVNGLYRHGFLAP 305

This analysis suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 97

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID

50 819>:

	1	ATGACTGATA	ATCGGGGGTT	TACGCTGGTT	GAATTAATAT	CAGTGGTCTT
	51	GATATTGTCT	GTACTTGCTT	TAATTGTTTA	TCCGAGCTAT	CGCAATTATG
	101	TTGAGAAAGC	AAAGATAAAT	GCACTGCGGG	CAGCCTTGTT	AGAAAATGCA
55	151	CATTTTATGG	AAAAGTTTTA	TCTGCAGAAT	GGGAGGTTTA	AACAAACATC
	201	TACCAAGTGG	CCAAGTTTGC	CGATTAAAGA	GGCAGAAGGC	TTTTGTATCC
	251	GTTTGAATGG	AATCGTCGCG	CGGG..GCTT	TAGACAGTAA	ATTCATGTTG
	301	AAGGCGGTAG	CCATAGATAA	AGATAAAAT	CCTTTTATTA	TTAAGATGAA
	351	TGAAAATCTA	GTAACCTTTA	ATTGCAAGA	AGTCCGCCAG	TTCGTGTAGT
	401	GACGGGCTGG	ATTATTTTAA	AGGAAATGAT	AAGGACTGCA	AGTTACTTAA
60	451	GTAG				

This corresponds to the amino acid sequence <SEQ ID 820; ORF127>:

```

      1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN
    51 HFMEKFYLQN GRFKQSTKW PSLPIKEAEG FCIRLNGIVA RXALDSKFML
   101 KAVAIDKDKN PFIKMENEL VTFICKKSAS CSDGLDYFK GNDKCKLLK
    151 *
  
```

Further work revealed the following DNA sequence <SEQ ID 821>:

```

      1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
    51 GATATTGTCT GACTTGCTT TAATTGTTA TCCGAGCTAT CGCAATTATG
   101 TTGAGAAAGC AAAGATAAAT GCAGTGGGG CAGCCTTGT AGAAAATGCA
   151 CATTTTATGG AAAAGTTTGA TCTGCAGAA GGGAGGTTA AACAAACATC
   201 TACCAAGTGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTGTATCC
   251 GTTTGAATGG AATCGCGCGC GGGGCTTAG ACAGTAAATT CATGTTGAAG
   301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
   351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAAGTACG
   401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG
  
```

This corresponds to the amino acid sequence <SEQ ID 822; ORF127-1>:

```

      1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN AVRAALLEN
    51 HFMEKFYLQN GRFKQSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
   101 AVAIDKDKNP FIIKMENELV TFICKKSASS CSDGLDYFKG NDKCKLLK*
  
```

20 Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF127 shows 98.0% identity over a 150aa overlap with an ORF (ORF127a) from strain A of *N. meningitidis*:

		10	20	30	40	50	60
25	orf127.pep	MTDNRGFTLV	ELISVVLILS	VLALIVYPSY	RNYVEKAKIN	AVRAALLEN	HFMEKFYLQN
	orf127a	MTDNRGFTLV	ELISVVLILS	VLALIVYPSY	RNYVEKAKIN	TVRAALLEN	HFMEKFYLQN
		10	20	30	40	50	60
30	orf127.pep	GRFKQSTKW	PSLPIKEAEG	FCIRLNGIAR	FXALDSKFML	KAVAIDKDKN	PFIKMENEL
	orf127a	GRFKQSTKW	PSLPIKEAEG	FCIRLNGI	ARGALDSKFM	KAVAIDKDKN	PFIKMENEL
		70	80	90	100	110	120
35	orf127.pep	VTFICKKSASS	CSDGLDYFKG	GNDKCKLLKX			
	orf127a	VTFICKKSASS	CSDGLDYFKG	GNDKCKLLKX			
40		120	130	140	150		

The complete length ORF127a nucleotide sequence <SEQ ID 823> is:

```

      1 ATGACTGATA ATCGGGGGTT TACGCTGGTT GAATTAATAT CAGTGGTCTT
    51 GATATTGTCT GACTTGCTT TAATTGTTA TCCGAGCTAT CGCAATTATG
   101 TTGAGAAAGC AAAGATAAAT ACAGTGGGG CAGCCTTGT AGAAAATGCA
   151 CATTTTATGG AAAAGTTTGA TCTGCAGAA GGGAGGTTA AACAAACATC
   201 TACCAATGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTGTATCC
   251 GTTTGAATGG AATCGCGCGC GGGGCTTAG ACAGTAAATT CATGTTGAAG
   301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
   351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTCG TGTAAGTACG
   401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG
  
```

This encodes a protein having amino acid sequence <SEQ ID 824>:

```

      1 MTDNRGFTLV ELISVVLILS VLALIVYPSY RNYVEKAKIN TVRAALLEN
    51 HFMEKFYLQN GRFKQSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
   101 AVAIDKDKNP FIIKMENELV TFICKKSASS CSDGLDYFKG NDKCKLLK*
  
```

-452-

ORF127a and ORF127-1 show 99.3% identity in 149 aa overlap:

```

      10      20      30      40      50      60
orfl27a.pep MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINTVRAALLENNAHFMEKFYLQN
5 orfl27-1   MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN
      10      20      30      40      50      60

      70      80      90      100     110     120
orfl27a.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIKMNENLV
10 orfl27-1   GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIKMNENLV
      70      80      90      100     110     120

      130     140     150
orfl27a.pep TFICKKSASSCSDGLDYFKGNDKDKCKLLKX
15 orfl27-1   TFICKKSASSCSDGLDYFKGNDKDKCKLLKX
      130     140     150

```

20 Homology with a predicted ORF from *N.gonorrhoeae*

ORF127 shows 97.3% identity over a 150 aa overlap with a predicted ORF (ORF127ng) from

N.gonorrhoeae:

```

orfl27.pep MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN 60
25 orfl27ng MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAAFLLENNAHFMEKFYLQN 60

orfl27.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIKMNENLV 120
30 orfl27ng GRFKQTSTKWPSLPIKEAEGFCIRLNGI-ARGALDSKFMLKAVAIDKDKNPFIKMNENLV 119

orfl27.pep VTTFICKKSASSCSDGLDYFKGNDKDKCKLLK 150
orfl27ng VTTFICKKSASSCSDRLDYFKGNDKDKCKLLK 149

```

The complete length ORF127ng nucleotide sequence <SEQ ID 825> is:

```

35 1 ATGACTGATA ATCGGGGGTT TACACTGGTT GAATTAATAT CAGTGGTCTT
51 GATATTGTCT GACTTGTCTT TAATTGTTTA TCCGAGCTAT CGCAATTATG
101 TTGAGAAAGC AAAGATAAAT GCAGTGCGGG CAGCCTTGTT AGAAAATGCA
151 CATTTTATGG AAAAGTTTGA TCTGCAGAAAT GGGAGATTGA AACAAACATC
201 TACCAAATGG CCAAGTTTGC CGATTAAAGA GGCAGAAGGC TTTTGATATC
40 251 GTTTGAATGG AATCGCGCGC GGGGCTTTAG ACAGTAAATT CATGTTGAAG
301 GCGGTAGCCA TAGATAAAGA TAAAAATCCT TTTATTATTA AGATGAATGA
351 AAATCTAGTA ACCTTTATTT GCAAGAAGTC CGCCAGTTTC TGTAGTGACG
401 GGCTGGATTA TTTTAAAGGA AATGATAAGG ACTGCAAGTT ACTTAAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 826>:

```

45 1 MTDNRGFTLV ELISVVLILSVLALIVYPSY RNYVEKAKIN AVRAAFLNA
51 HMEKFYLQN GRFKQTSTKW PSLPIKEAEG FCIRLNGIAR GALDSKFMLK
101 AVAIDKDKNP FIIKMNENLV TFICKKSASS CSDRLDYFKG NDKDCKLLK*

```

ORF127ng and ORF127-1 show 100.0% identity in 149 aa overlap:

```

      10      20      30      40      50      60
50 orfl27-1.pep MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN
orfl27ng-1 MTDNRGFTLV ELISVVLILSVLALIVYPSYRNYVEKAKINAVRAALLENNAHFMEKFYLQN
      10      20      30      40      50      60

      70      80      90      100     110     120
55 orfl27-1.pep GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIKMNENLV
orfl27ng-1 GRFKQTSTKWPSLPIKEAEGFCIRLNGIARGALDSKFMLKAVAIDKDKNPFIKMNENLV
      70      80      90      100     110     120
60

```

-453-

```

              130      140      150
orf127-1.pep  TFICKKSASSCS DGLDYFKGNDKDKLLKX
              |||
orf127ng-1    TFICKKSASSCS DGLDYFKGNDKDKLLKX
              130      140      150

```

This analysis, including the fact that the predicted transmembrane domain is shared by the meningococcal and gonococcal proteins, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

10 Example 98

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 827>

```

      1  ..GTGTCGCTGG CTTGGGTGAT TGCCTCTCAA ATCTTCCTTT ACGAAGATTT
     51  CAACCAAATG CCGAAAAACCG GTGGAGCTAT CTGCGGTTTT CTTGTCCAAT
    101  ATTTATCTGG GGTTCAGCA GGGGTATTTG GATTTGAGTG CCGACGAGAA
    151  CCCCGTACTG CATATCTGGT CTTTGCCAGT AGAGGAACAG TATTACCTCC
    201  TGTATCCCTT TTTGCTGATA TTTTGCTGCA AAAAAACCAA ATCGCTACGG
    251  GTGCTGCGTA ACATCAGCAT CATCCTGTTT TTGATTTTGA CTGCCTCATC
    301  GTTTTTGCCA AGCGGGTTTT ATACCGACAT CCTCAACCAA CCCAATACTT
    351  ATTACCTTTC GACACTGAGG TTTCCCGAGC TGTGGCAGG TTCGCTGCTG
    401  GCGGTTTACG GGCAAAACGCA AAACGGCAGA CGGCAAAACAG CAAATGGAAA
    451  ACGGCAGTTG CTTTCATCAC TCTGCTTCGG CGCATTGCTT GCCTGCCTGT
    501  TCGTGATTGA CAAACACAAT CCGTTTATCC CGGGAATGAC CCTGCTCCTT
    551  CCCTGCCTGC TGACGGCACT GCTTATCCGG AGTATGCAAT ACGGGACACT
    601  TCCGACCCGC ATCCTGTCCG CAAGCCCCAT CGTATTTGTC GGCAAAATCT
    651  CTTATTCCTT ATACCTGTAC CATGGGATT TTATTGCTTT CGCTCCGCTC
    701  ATTAGAGGCG GGAACAGCT CGGACTGCCT GCCG..

```

This corresponds to the amino acid sequence <SEQ ID 828; ORF128>:

```

      1  ..VSLASVIASQ IFLYEDFNQM RKTVELSAVF LSNIYLGFOQ GYFDLSADEN
     51  PVLHIWSLAV EEQYLYLLPL LLIFCKKTK SLRVLNRNISI ILFLILTASS
    101  FLPSGFYTDI LNQPNYYLS TLRPELLAG SLLAVYGQTO NGRRTANGK
    151  RQLSSLCFG ALLACLFVID KHNPFIPGMT LLLPCLLTAL LIRSMQYGTL
    201  PTRILSASPI VEVGKISYSL YLYHWIFIAF APLIRGGKQL GLPA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 829>:

```

      1  ATGCAAGCTG TCCGATACAG ACCGGAAATT GACGGATTGC GGGCCGTCGC
     51  CGTGCTATCC GTCATGATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG
    101  GATTCCTGGG GGTGGACATT TTCTTTGTCA TCTCAGGATT CCTCATTACC
    151  GGCATCATTC TTTCTGAAAT ACAGAACGGT TCTTTTCTT TCCGGGATTT
    201  TTATACCCGC AGGATTAAGC GGATTTATCC TGCCTTTATT GCGGCCGTGT
    251  CGCTGGCTTC GGTGATTGCC TCTCAAATCT TCCTTTACGA AGATTTC AAC
    301  CAAATGCGGA AAACCGTGGA GCTTCTGCG GTTTCTTGT CCAATATTTA
    351  TCTGGGGTTT CAGCAGGGGT ATTTCGATTT GAGTGCCGAC GAGAACCCCG
    401  TACTGCATAT CTGGTCTTTG GCAGTAGAGG AACAGTATTA CCTCCTGTAT
    451  CCCCTTTTGC TGATATTTTG CTGCAAAAAA ACCAAATCGC TACGGGTGCT
    501  GCGTAACATC AGCATCATCC TGTTTTGTAT TTTGACTGCC TCATCGTTTT
    551  TGCCAAGCGG GTTTTATACC GACATCCTCA ACCAACCCAA TACTTATTAC
    601  CTTTCGACAC TGAGGTTTCC CGAGCTGTTG GCAGGTTTCG TGCTGGCGGT
    651  TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGCAAAAT GGAAACCGG
    701  AGTTGCTTTC ATCACTCTGC TTCGGCGCAT TGCTTGCTG CCTGTTCTGT
    751  ATTGACAAAC ACAATCCGTT TATCCCGGGA ATGACCCTGC TCCTTCCCTG
    801  CCTGCTGACG GCACTGCTTA TCCGGAGTAT GCAATACGGG ACACTTCCGA
    851  CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT
    901  TCCCTATACC TGTACCATTG GATTTTATT GCTTTCGCCC ATTACATTAC
    951  AGGCGACAAA CAGCTCGGAC TGCCTGCCGT ATCGGCGGTT GCCGCGTTGA
   1001  CGGCCGGATT TTCCCTGTTG AGTTATTATT TGATTGAACA GCCGCTTAGA
   1051  AAACGGAAGA TGACCTTCAA AAAGGCATT TTCTGCCTCT ATCTCGCCCC
   1101  GTCCCTGATA CTTGTCGGTT ACAACCTGTA CGCAAGGGGG ATATTGAAAC
   1151  AGGAACACCT CCGCCCGTTG CCCGGCGCGC CCCTTGCTGC GGAAATCAT

```

-454-

5
10
15
20
25

```

1201 TTTCCGGAAA CCGTCTGAC CCTCGGCGAC TCGCAGCCG GACACCTGAG
1251 GGGGTTTCTG GATTATGTCG GCAGCCGGA AGGGTGGAAA GCCAAATCC
1301 TGTCCTCTGA TTCGGAGTGT TTGGTTTGGG TAGATGAGAA GCTGGCAGAC
1351 AACCCGTTAT GTCGAAAATA CCGGGATGAA GTTGAAAAAG CCGAAGCCGT
1401 TTTCAATTGCC CAATTCTATG ATTTGAGGAT GGGCGGCCAG CCTGTGCCGA
1451 GATTTGAAGC GCAATCCTTC CTAATACCCG GGTTCACAGC CCGATTACAG
1501 GAAACCGTCA AAAGGATAGC CGCCGTCAA CCCGTCTATG TTTTGCAAA
1551 CAACACATCA ATCAGCCGTT CGCCCTGAG GGAGGAAAAA TTGAAAAGAT
1601 TTGCCGCAAA CCAATATCTC CGCCCATTC AGGCTATGGG CGACATCGGC
1651 AAGAGCAATC AGGCGGTCTT TGATTTGATT AAAGATATTC CCAATGTGCA
1701 TTGGGTGGAC GCACAAAAT ACCTGCCCAA AAACACGGTC GAAATATACG
1751 GCCGCTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT
1801 TATATGGGGC GGAATTCAC CAAACACGAA CGCCTGCTTA AATCTTCCCA
1851 CGGCGGCGCA TTGCAGTAG

```

15 This corresponds to the amino acid sequence <SEQ ID 830; ORF128-1>:

20
25

```

1 MQAVRYRPEI DGLRAVAVLS VMIFHLNNRW LPGGFLGVDI FFVISGFLIT
51 GIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN
101 QMRKTVELSA VFSLNIYLG FQQGYFDLSAD ENPVLHIWSL AVEEQYLLY
151 PLLLI FCKK TKSLRVLNRI SIILFLILTA SSFLPSGFYT DILNQPNYY
201 LSTLRFPELL AGSLLAVYGQ TQNGRRQTAN GKROLLSSLC FGALLACLFV
251 IDKHNPFIPG MTL LLPCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY
301 SLYLYHWIFI AFAHYITGDK QLG LPAVSAV AALTAGFSL SYLLIEQPLR
351 KRKMTFKKAF FCLYLAPSLI LVGYNLYARG ILKQEHRLPL PGAPLAAENH
401 FPETVLTLDG SHAGHLRGFL DYVGSREGWK AKILSLDSEC LVWVDEKLAD
451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPRFEAQSF LIPGFPARFR
501 ETVKRIAAVK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG
551 KSNQAVFDLI KDIPNVHWVD AQKYL PKNTV EIYGRYLYGD QDHLTYFGSY
601 YMGREFHKHE RLLKSSHGGA LQ*

```

Computer analysis of this amino acid sequence gave the following results:

30 Homology with hypothetical integral membrane protein HI0392 of *H. influenzae* (accession number U32723)
ORF128 and HI0392 show 52% aa identity in 180aa overlap:

35
40

```

Orf128: 1 VSLASVIASQIFLYEDFNQMRKTVELSAVFSLNIYLG FQQGYFDLSADENPVLHIWSLAV 60
++L S IAS IF+Y DEN++RKT+EL+ FLSN YLG QGYFDLSA+ENPVLHIWSLAV
HI0392: 46 MALVSFIASAIIFYNDENKLRKTIELAI AFLSNFYLG LQTGYFDLSANENPVLHIWSLAV 105

Orf128: 61 EEQXXXXXXXXXIFCKKTKSLRVLNRI SIILFLILTASSFLPSGFYT DILNQPNYYLS 120
E Q I KK + ++VL I++ILF IL A+SF+ + FY ++L+QPN YYLS
HI0392: 106 EGQYYLIFPLILILAYKKFREVKVLFITLILFFILLATSFVSANFYKEVLHQPNIYYLS 165

Orf128: 121 TLRFPPELLAGSLLAVYGQTQNGRRQTANGKROLLSSLCFGALLACLFVIDKHNPFIPGMT 180
LRFPPELL GSLLA+Y N + Q + +L+ L L +CLF+++ + FIPG+T
HI0392: 166 NLRFPPELLVGSLLAIYHNLSN-KVQLSKQVNNILAILSTLLFSCFLMNNNIAFIPGIT 224

```

Homology with a predicted ORF from *N. meningitidis* (strain A)

45 ORF128 shows 98.0% identity over a 244aa overlap with an ORF (ORF128a) from strain A of *N. meningitidis*:

50
55
60

```

          10      20      30
orf128.pep          VSLASVIASQIFLYEDFNQMRKTVELSAVF
                      |||
orf128a      ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVF
              60      70      80      90      100     110

          40      50      60      70      80      90
orf128.pep      LSNIYLG FQQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNRI SI
              120     130     140     150     160     170
orf128a      LSNIYLG FQQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNRI SI
              120     130     140     150     160     170

          100     110     120     130     140     150
orf128.pep      ILFLILTASSFLPSGFYT DILNQPNYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGK
              |||

```

-455-

```

orfl28a      ILEFLILTATSF LPSGFYTDILNQPNNTYYLSTLRFPPELLAGSLLAVYGQTONGRRQTANGK
              180      190      200      210      220      230

5  orfl28.pep RQLSSSLCFGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI
              160      170      180      190      200      210
orfl28a      RQLSSSLCFGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI
              240      250      260      270      280      290

10 orfl28.pep VFVGKISYSLYLHWIFIAFAPLIRGGKQLGLPA
              220      230      240
orfl28a      VFVGKISYSLYLHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKR
              300      310      320      330      340      350

15 orfl28a      KMTFFKAFFCLYLAPSLILVGYNLYARGILKQEHLRPLPGAPLAAENHFPETVLTGLDSH
              360      370      380      390      400      410

```

The complete length ORF128a nucleotide sequence <SEQ ID 831> is:

```

20 1 ATGCAAGCTG TCCGATACAG ACCGGAATTT GACGGATTGC GGGCCGTCGC
51 CGTGCTATCC GTCATGATTT TCCACCTGAA TAACCGCTGG CTGCCCGGAG
101 GATTCTCTGGG GGTGGACATT TTCTTTGTCA TCTCAGGATT CCTCATTACC
151 GGCATCATTCT TTTCTGAAAT ACAGAACGGT TCTTTTCTTT TCCGGGATTT
201 TTATACCCGC AGGATTAAGC GGATTTATCC TGCTTTTATT GCGGCCGTGT
251 CGCTGGCTTC GGTGATTGCC TCTCAAATCT TCCTTTACGA AGATTTCAAC
301 CAAATGCGGA AAACCGTGGA GCTTTCTGCG GTTTTCTGTG CCAATATTTA
351 TCTGGGGTTT CAGCAGGGGT ATTTCGATT GAGTGGCCGAC GAGAACCCCG
401 TACTGCATAT CTGGTCTTTG GCAGTAGAGG AACAGTATTA CCTCCTGTAT
451 CCTCTTTTGC TGATATTTTG CTGCAAAAAA ACAAAATCGC TACGGGTGCT
501 CGGTAACATC AGCATCATCC TATTTCTGAT TTTGACTGCC ACATCGTTTT
30 551 TGCCAAGCGG GTTTTATACC GATATTCTCA ACCAACCCAA TACTTATTAC
601 TTTTCGACAC TGAGGTTTCC CGAGCTGTTG GCAGGTTCGC TGCTGGCGGT
651 TTACGGGCAA ACGCAAAACG GCAGACGGCA AACAGCAAAT GGAAAACGGC
701 AGTTGCTTTC ATCACTCTGC TTCGGCGCAT TGCTTGCCTG CCTGTTCTGT
751 ATTGACAAAC ACAATCCGTT TATCCCGGGA ATGACCCTGC TCCTTCCCTG
35 801 CCTGCTGACG GCACTGCTTA TCCGGAGTAT GCAATACGGG ACACCTCCGA
851 CCCGCATCCT GTCGGCAAGC CCCATCGTAT TTGTCGGCAA AATCTCTTAT
901 TCCCTATACC TGTACCATTG GATTTTATT GCTTTCGCCC ATTACATTAC
951 AGGCGACAAA CAGCTCGGAC TGCCTGCCGT ATCGGCGGTT GCCGCGTTGA
1001 CGGCCGGATT TTCCCTGTG AGTTATTATT TGATTGAACA GCCGCTTAGA
40 1051 AAACGGAAGA TGACCTTCAA AAAGGCATT TTCTGCCTCT ATCTCGCCCC
1101 GTCCCTGATA CTTGTGCGTT ACAACCTGTA CGCAAGGGGG ATATTGAAAC
1151 AGGAACACCT CCGCCCGTTG CCGGCGCGCG CCCTTGCTGC GGAAAATCAT
1201 TTTCCGGAAG CCGTCTGAC CCTCGGCGAC TCGCACGCCG GACACCTGCG
1251 GGGGTTTCTG GATTATGTCG GCAGCCGGGA AGGGTGGAAG GCCAAAATCC
45 1301 TGTCCCTCGA TTCGGAGTGT TTGGTTTGGG TAGATGAGAA GCTGGCAGAC
1351 AACCCTGTTAT GTCGAAAATA CCGGGATGAA GTTGAAAAAG CCGAAGCCGT
1401 TTTTATTGCC CAATCTATG ATTTGAGGAT GGGCGGCCAG CCCGTGCCGA
1451 GATTGAAGC GCAATCCTTC CTAATACCCG GGTTCACGAG CCGATTGAGG
50 1501 GAAACCGTCA AAAGGATAGC CGCCGTCAA CCCGTCTATG TTTTTCGAAA
1551 CAACACATCA ATCAGCCGTT CGCCCTGAG GGAGGAAAAA TTGAAAAGAT
1601 TTGCCGCAAA CCAATATCTC CGCCCATTC AGGCTATGGG CGACATCGGC
1651 AAGAGCAATC AGGCGGTCTT TGATTGATT AAAGATATTC CCAATGTGCA
1701 TTGGGTGAC GCACAAAAAT ACCTGCCCAA AAACACGGTC GAAATATACG
55 1751 GCCGCTATCT TTACGGCGAC CAAGACCACC TGACCTATTT CGGTTCTTAT
1801 TATATGGGCG GGGAATTTC CAAACACGAA CGCCTGCTTA AATCTTCTCG
1851 CGACGGCGCA TTGCAGTAG

```

This encodes a protein having amino acid sequence <SEQ ID 832>:

```

60 1 MQAVRYRPEI DGLRAVAVLS VMIFHLNNRW LPGGFLGVDI FFVISGFLIT
51 GIILSEIQNG SFSFRDFYTR RIKRIYPAFI AAVSLASVIA SQIFLYEDFN
101 QMRKTVELSA VFLSNIYLG FQQGYFDLSAD ENPVLHIWSL AVEEQYLLY
151 PLLLIFFCKK TKSLRVLRLNI SIILFLILTA TSFLPSGFYT DIILNQPNNTYY
201 LSTLRFPPELL AGSLLAVYGO TONGRRQTAN GKRQLSSSLC FGALLACLFV
251 IDKHNPFIPG MTL LLPCLLT ALLIRSMQYG TLPTRILSAS PIVFVGKISY
301 SLYLYHWIFI AFAHYITGDK QLGLPAVSAV AALTAGFSLL SYYLIEQPLR
65 351 KRKMTFFKAF FCLYLAPSLI LVGYNLYARG ILKQEHLRPL PGAPLAAENH
401 FPETVLTGLD SHAGHLRGFL DYVGSREGWK AKILSLDSEC LVWVDEKLAD
451 NPLCRKYRDE VEKAEAVFIA QFYDLRMGGQ PVPRFEAQSF LIPGFPARFR

```

-456-

501 ETVKRIAANK PVYVFANNTS ISRSPLREEK LKRFAANQYL RPIQAMGDIG
 551 KSNQAVFDLI KDIPNVHWVD AQKYLPKNTV EIYGRYLYGD QDHLTYFGSY
 601 YMGREFHKHE RLLKSSRDGA LQ*

ORF128a and ORF128-1 show 99.5% identity in 622 aa overlap:

5	orf128a.pep	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
	orf128-1	MQAVRYRPEIDGLRAVAVLSVMIFHLNNRWLPGGFLGVDIFFVISGFLITGIILSEIQNG
10	orf128a.pep	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
	orf128-1	SFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTVELSAVFLSNIYLG
	orf128a.pep	QQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNISIILFLILTA
15	orf128-1	QQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNISIILFLILTA
	orf128a.pep	TSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKROLLSSLC
	orf128-1	SSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGKROLLSSLC
20	orf128a.pep	FGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY
	orf128-1	FGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASPIVFGKISY
25	orf128a.pep	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLSYLLIEQPLRKRMFTFKKAF
	orf128-1	SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLSYLLIEQPLRKRMFTFKKAF
30	orf128a.pep	FCLYLAPSLILVGNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
	orf128-1	FCLYLAPSLILVGNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLGDSHAGHLRGFL
	orf128a.pep	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
35	orf128-1	DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
	orf128a.pep	PVPRFEAQSFILPGFPPARFRET VKRIAANKPVYVFANNTSISRSPLREEKLRFAANQYL
	orf128-1	PVPRFEAQSFILPGFPPARFRET VKRIAANKPVYVFANNTSISRSPLREEKLRFAANQYL
40	orf128a.pep	RPIQAMGDIGKSNQAVFDLIKDI PNHWVDAQKYLPKNTVEIYGRYLYGDQDHLTYFGSY
	orf128-1	RPIQAMGDIGKSNQAVFDLIKDI PNHWVDAQKYLPKNTVEIYGRYLYGDQDHLTYFGSY
45	orf128a.pep	YMGREFHKHERLLKSSRDGALQX
	orf128-1	YMGREFHKHERLLKSSHGALQX

Homology with a predicted ORF from *N.gonorrhoeae*

50 ORF128 shows 93.4% identity over 244 aa overlap with a predicted ORF (ORF128ng) from *N. gonorrhoeae*:

	orf128.pep	VSLASVIASQIFLYEDFNQMRKTVELSAVF	30
	orf128ng	ILSEIQNGSFSFRDFYTRRIKRIYPAFIAAVSLASVIASQIFLYEDFNQMRKTIELSTVF	112
55	orf128.pep	LSNIYLGFGQGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCCKKTKSLRVLNISI	90
	orf128ng	LSNIYLGFR LGYFDLSADENPVLHIWSLAVEEQYLLYPLLLIFCYKTKTKSLRVLNISI	172
60	orf128.pep	ILFLILTASSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLLAVYGQTQNGRRQTANGK	150
	orf128ng	ILFLILTASSFLPAGFYTDILNQPNYYLSTLRFPPELLVGSLLAVYGQTQNGRRQTENGK	232
	orf128.pep	RQLLSSLCFGALLACLFVIDKHNPFI PGMTLLLPCLLTALLIRSMQYGTLPTRILSASPI	210
65	orf128ng	RQLLSSLCFGALLVCLFVIDKHNPFI PGITLLLPCLLTALLIRSMQYGTLPTRILSASPI	292

	1	ATGCAAGCTG	TCCGATACAG	GCCTGAAATT	GACGGATTGC	GGGCCGTCGC
	51	CGTGCTATCC	GTCATTATTT	TCCACCTGAA	TAACCGCTGG	CTGCCCGGAG
	101	GATTCCCTGGG	GGTGACATT	TTCTTTGTCA	TCTCGGGATT	CCTCATTACC
10	151	AACATCATT	TTTCTGAAAT	ACAGAACGTT	TCCTTTTCTT	TCCGGGATTT
	201	TTATACCCGC	AGGATTAAGC	GGATTTATCC	TGCTTTTATT	GCGGCCGTGT
	251	CCCTGGCTTC	GGTGATTGCT	TCTCAAATCT	TCCTTTACGA	AGATTTCAAC
	301	CAAATGAGGA	AAACCATAGA	GCTTTCTACG	GTTTTTTTGT	CCAATATTTA
	351	TTTGGGGTTC	CGATTGGGGT	ATTCGATT	GAGTGCCGAC	GAGAACCCCG
	401	TACTGCATAT	CTGGTCTTTG	GCGGTAGAGG	AACAGTATTA	CCTCCTGTAT
15	451	CCTCTTTTGC	TGATATTCTG	TTACAAAAAA	ACCAATATC	TACGGGTGCT
	501	GGCTAATATC	AGCATCATCC	TGTTTCTGAT	TTTGACCGCA	TCATCGTTT
	551	TGCCGGCCGG	GTTTATAC	GACATCCTCA	ACCAACCCaa	TACTTATTAC
	601	CTTTCGACAC	TGAGGTTTCC	CGAGCTGTTG	GTGGGTTCGC	TGTTGGCCGT
	651	TTACGGGGCAA	ACGCAAAACG	GCAGACGGCA	AACAGAAAAAT	GGAAAAACCG
20	701	AGTTGCTTTC	ATTACTCTGT	TTCGGCGCat	tgCTTGTCGT	CCTGTTCCGT
	751	ATCGACAAC	ACGATCCGTT	TATCCCGGGA	ATAACCTG	TCCTTCCCTG
	801	CCTGCTGACG	GCGTGTCTTA	TCCGGAGTAT	GCAATACGGG	ACACTTCCGA
	851	CCCGCATCCT	GTCGGCAAGC	CCCATCGTAT	TTGTCGGCAA	AATCTCTTAT
	901	TCCCTATAC	TGTACCATTG	GATTTTATT	GCCTTCGCCC	ATTACATTAC
25	951	AGGCGCAAAA	CAGCTCGGAC	TGCCCTGCCG	ATCGGCGGTT	GCCGCGTTGA
	1001	CGGCCGGATT	TTCCCTGTTG	AGCTATTATT	TGATTGAACA	GCCGCTTAGA
	1051	AAACGGAAGA	TGACCTTCAA	AAAGGCATT	TTCTGCCTTT	ATCTCGCCCC
	1101	TGCCCTGATG	CTTGTGCGTT	ACAACCTGTA	TTCAAGAGGG	ATATTGAAAC
	1151	AGGAACACCT	CCGCCCGCTG	CCCGGCACGC	CCGTTGCTGC	GGAAAAAAT
30	1201	TTTCCGGAAA	CCGCTCTGAC	CCTCGGCGAC	TGCGACGCCG	GACACCTGCG
	1251	GGGGTTTCTG	GATTATGTCT	GCGGCAGGGA	AGGGTGAAAA	GCTAAAAATC
	1301	TGTCCTCGA	TTCCGAGTGT	TTGGTTGGG	TGGATGAGAA	GCTGGCAGAC
	1351	AACCCGTTGT	GCGGAAAAAT	CCGGGATGAA	GTTGAAAAAG	CCGAAGCTGT
	1401	TTTCATTGCC	CAATTCTATG	ATTTAGGAT	GGGCGGCCAG	CCCGTGCCGA
35	1451	GATTTGAAGC	GCAATCCTTC	CTGATACCCG	GCTTCAAAGC	CCGATTCAAG
	1501	GAAACCGTCA	AGAGGATAGC	CGCCGTCAAA	CCTGTATATG	TTTTTGCAAA
	1551	CAATACATCA	ATCAGCGCTT	CTCCCTTGAG	GGAGGAAAAA	TTGAAAAAGT
	1601	TTGCTATAAA	CCAATACCTC	CGGCCTATT	GGGCTATGGG	CGACATCGGC
	1651	AAGAGCAAT	AGGCGGTCTT	TGATTTGGTT	AAAGATATTA	CCAATGTGCA
40	1701	TGGGGTGGAC	GCACAAAAAT	ACCTGCCCAA	AAACACGGTC	GAAATACACG
	1751	GACGCTATCT	TTACGGCGAC	CAAGACCACC	TGACCTATT	CGGTTCTTAT
	1801	TATATGGGGG	GGGAATTCCA	CAACACGAA	CGCCTGCTCA	AGCATTTCCC
	1851	AGCGGGCGCA	TTGCAATAG			

45	1	<u>MQAVRYRPEI</u>	<u>DGLRAVAVLS</u>	<u>VIIFHLNNRW</u>	<u>LPGGFLGVDI</u>	<u>FFVISGFLIT</u>
	51	<u>NIILSEIQNG</u>	<u>SFSFRDFYTR</u>	<u>RIKRIYPAFI</u>	<u>AAVSLASVIA</u>	<u>SQIFLYEDFN</u>
	101	<u>QMRKTIELST</u>	<u>VFLSNIIYLG</u>	<u>RLGYFDLSAD</u>	<u>ENPVLIHWSL</u>	<u>AVEEQYLLY</u>
	151	<u>PLLLIFCYKK</u>	<u>TKSLRVLRLNI</u>	<u>SIILFLILTA</u>	<u>SNVFLPAGFYT</u>	<u>DILNQNTYY</u>
50	201	<u>LSTLRFPPELL</u>	<u>VGSLLAYVYG</u>	<u>TQNGRRQTEN</u>	<u>GKRQLLSLLC</u>	<u>FGALLVCLFV</u>
	251	<u>IDKHDPFIPG</u>	<u>ITLLLPCLLT</u>	<u>ALLIRSMQYG</u>	<u>TLPTRIILAS</u>	<u>PIVFGKISY</u>
	301	<u>SLYLYHWIFI</u>	<u>AFAYYITGDK</u>	<u>QLGLPAVSAV</u>	<u>AALTAGFSLL</u>	<u>SYTIEQPLR</u>
	351	<u>KRKMTFKKAF</u>	<u>FCLYLAPSLM</u>	<u>LVGYNLYSRG</u>	<u>ILKQEHRLPL</u>	<u>PGTPVAEANN</u>
55	401	<u>FPETVLTLGD</u>	<u>SHAGHLRGFL</u>	<u>DYVGGREGWK</u>	<u>AKILSLDSEC</u>	<u>LWVWDEKLAD</u>
	451	<u>NPLCRKYRDE</u>	<u>VEKAEAVFIA</u>	<u>QFYDLRMGGQ</u>	<u>PVPFEAAQSF</u>	<u>LIPGFKARFR</u>
	501	<u>ETVKRIAIAV</u>	<u>PVYVFANNTS</u>	<u>ISRSPREEK</u>	<u>LKRFAINQYL</u>	<u>RPIRAMGDIG</u>
	551	<u>KSNQAVFDLV</u>	<u>KDIPNVHWVD</u>	<u>AQKYLPKNTV</u>	<u>EIHGRYLYGD</u>	<u>QDHLTYFGSG</u>
	601	<u>YMGREFHKHE</u>	<u>RLCHKSRGGA</u>	<u>LQ*</u>		

[illegible]


```

5  orf128-1.pep  OQGYFDLSADENPVLHIWSLAVEEQYLLYP LLLIFCCKKTKSLRVLNRNISIILFLILTA
   orf128ng      RLGYFDLSADENPVLHIWSLAVEEQYLLYP LLLIFCYKKTSLRVLNRNISIILFLILTA
   orf128-1.pep  SSFLPSGFYTDILNQPNYYLSTLRFPPELLAGSLLAVYGQTONGRRQTANGKRQLLSLSC
   orf128ng      SSFLPAGFYTDILNQPNYYLSTLRFPPELLVGSLLAVYGQTONGRRQTENGKRQLLSLSC
10  orf128-1.pep  FGALLACLFVIDKHNPFIPGMTLLLPCLLTALLIRSMQYGTLPTRILSASP I VFGKISY
   orf128ng      FGALLVCLFVIDKHDPFIPGITLLLPCLLTALLIRSMQYGTLPTRILSASP I VFGKISY
15  orf128-1.pep  SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
   orf128ng      SLYLYHWIFIAFAHYITGDKQLGLPAVSAVAALTAGFSLLSYYLIEQPLRKRKMTFKKAF
20  orf128-1.pep  FCLYLAPSLIIVGYNLYARGILKQEHRLPLPGAPLAAENHFPETVLTGLDSSHAGHLRGFL
   orf128ng      FCLYLAPSLMLVGYNLYSRGILKQEHRLPLPGTPVAAENHFPETVLTGLDSSHAGHLRGFL
25  orf128-1.pep  DYVGSREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
   orf128ng      DYVGGREGWKAKILSLDSECLVWVDEKLADNPLCRKYRDEVEKAEAVFIAQFYDLRMGGQ
30  orf128-1.pep  PVPRFEAQSFILPGFPARFRET VKRIA AVKPVYVFANNTSISRSP LREEKLKRFAANQYL
   orf128ng      PVPRFEAQSFILPGFKARFRET VKRIA AVKPVYVFANNTSISRSP LREEKLKRFAANQYL
35  orf128-1.pep  RPIQAMGDIGKSNQAVFDLIK DIPNVHWVDAQYLPKNTVEIYGRYLYGDQDHLTYFGSY
   orf128ng      RPIRAMGDIGKSNQAVFDLVKDIPNVHWVDAQYLPKNTVEIHGRYLYGDQDHLTYFGSY
   orf128-1.pep  YMGREFHKHERLLKSSHGALQX
   orf128ng      YMGREFHKHERLLKHSRGALQX
                        610      620

```

In addition, ORF218ng shows homology to a hypothetical *H. influenzae* protein:

```

40  sp|P43993|Y392_HAEIN  HYPOTHETICAL PROTEIN HI0392 >gi|1074385|pir|B64007
   hypothetical protein HI0392 - Haemophilus influenzae (strain Rd KW20)
   >gi|1573364 (U32723) H. influenzae predicted coding region HI0392 [Haemophilus
   influenzae] Length = 245
   Score = 239 bits (604), Expect = 3e-62
   Identities = 124/225 (55%), Positives = 152/225 (67%), Gaps = 1/225 (0%)
45  Query: 38  VDIFFVISGFLITNIILSEIQNGSFSFRDFYTRRIKRIYPXXXXXXXXXXXXXXXXXXFLYE 97
   +DIFFVISGFLIT II++EIQ SFS + FYTRRIKRIYP F+Y
   Sbjct: 1  MDIFFVISGFLITGIIITEIQNSFSLKQFYTRRIKRIYPAFITVMALVSFIASAIIFYN 60
50  Query: 98  DFNQMRKTIELSTVFLSNIIYLGFR LGYFDLSADENPVLHIWSLAVEEQXXXXXXXXXXIFC 157
   DFN++RKTIEL+ FLSN YLG GYFDLSA+ENPVLHIWSLAVE Q I
   Sbjct: 61  DFNKLRKTIELAIAFLSNFYLG LTQGYFDLSANENPVLHIWSLAVEGQYYLIFLILILA 120
55  Query: 158 YKKTSLRVLNRNISIILFLILTASSFLPAGFYTDILNQPNYYLSTLRFPPELLVGSLLAV 217
   YKK + ++VL I++ILF IL A+SF+ A FY ++L+QPN YYLS LRFPELLVGSLLA+
   Sbjct: 121 YKKFREVKVLFIITLILFFILLATSFVSANFYKEVLHQPNIIYLSNLRFPPELLVGSLLAI 180
60  Query: 218 YGQTONGRRQTENGKRQLLSLLCFGALLVCLFVIDKHDPFIPGIT 262
   Y N + Q +L++L L CLF+++ + FIPGIT
   Sbjct: 181 YHNLSN-KVQLSKQVNNILAILSTLLLFSCFLMNNNIAFIPGIT 224

```

This analysis, including the identification of several putative transmembrane domains, suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 99

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 835>:

```

1  ..ATTATTTACG AATACCGCTG GATGTTTCTT TACGGCGCAC TGACGACCTT
51  GGGGCTGACG GTCGTGGCAA C.GCGGGCGG TTCGGTATTG GGTCTGTTGT
101 TGGCGTTGGC GCGCCTGATT CACTTGGAAG AAGCCGGTGC GCCGATGCCG
151 GTGCTGGCGT GGGCGTTGCG TAAAGTTTCG CTGCTGTATG TTACGCTGTT
201 CCGGGGTACG CCGCTGTTTG TGCAGATTGT GATTGGGGCG TATGTGTGGT
251 TTCCGTTTTT CGTC..

```

This corresponds to the amino acid sequence <SEQ ID 836; ORF129>:

```

1  ..IIYEYRWMFL YGALTTLGLT VVAXAGGSVL GLLLALARLI HLEKAGAPMR
51  VLAWALRKVS LLYVTLFRGT PLFVQIWIWA YVWFPPFFV..

```

Further work revealed the complete nucleotide sequence <SEQ ID 837>:

```

1  ATGGATTTTC GTTTTGACAT TATTTACGAA TACCGCTGGA TGTTCCTTTA
51  CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCAACG GCGGGCGGTT
101 CCGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA
151 GCCGGTGCGC CGATCGCGGT GCTGGCGTGG GCGTTGCGTA AAGTTTCGCT
201 GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTGTG CAGATTGTGA
251 TTTGGGCGTA TGTGTGGTTT CCGTTTTCG TCCATCCTTC AGACGGCATT
301 TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
351 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
401 AGATTTTCCG CGCGGCGATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
451 GCGCGTTCTT TGGGGCTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT
501 GCCGCAGGCA TTGCGCCGCA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGGGAGTTG
601 GCGTATGTTT AGAATACGAT TACGGGCGCG TATTCGGTTT ATGAAGAACC
651 GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
701 GGATATTCCT GCGTTTGAA AAACGTTACA ATCCGCAACA CCGCTGA

```

This corresponds to the amino acid sequence <SEQ ID 838; ORF129-1>:

```

1  MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51  AGAPMRVLAW ALRKVSLLYV TLFRTPLFV QIWIWAYVWF PFFVHPSDGI
101 LVSGEAAIAL RRGYGLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA
151 ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL
201 AYWQNTITGR YSVYEEPLYT VALIYLLMT FLGWIFLRLE KRYNPQHR*

```

Computer analysis of this amino acid sequence gave the following results:

35 Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF129 shows 98.9% identity over a 88aa overlap with an ORF (ORF129a) from strain A of *N.meningitidis*:

```

10      20      30      40      50
40  orf129.pep      IIYEYRWMFLYGALTTLGLTVVAXAGGSVLGLLLALARLIHLEKAGAPMRVLAW
    orf129a      MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
10      20      30      40      50      60
45  orf129.pep      ALRKVSLLYVTLFRGTPLFVQIWIWAYVWFPPFFV
    orf129a      ALRKVSLLYVTLFRGTPLFVQIWIWAYVWFPPFFVHPSDGILVSGEAAIALRRGYGLIAG
70      80      90      100     110     120
50  orf129a      SLALIANSQA YICEIFRAGI QSIDKGQMEA ARSLGLTYPQ AMRYVILPQA LRRMLPLAS
130     140     150     160     170     180

```

The complete length ORF129a nucleotide sequence <SEQ ID 839> is:

```

1  ATGGATTTTC GTTTTGACAT TATTTACGAA TACCGCTGGA TGTTCCTTTA
51  CGGCGCACTG ACGACCTTGG GGCTGACGGT CGTGGCGACG GCGGGCGGTT

```

5
10
15
20
25
30
35
40
45
50
55
60
65
70

```

101 CGGTATTGGG TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTTGGAAAAA
151 GCCGGTGCGC CGATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT
201 GCTGTATGTT ACGCTGTTCC GGGGTACGCC GCTGTTTGTG CAGATTGTGA
251 TTTGGGCGTA TGTGTGGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT
301 TTGGTTAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
351 GATTGCCGCT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAGGACA GATGGAGGCG
451 GCGCGTTCTT TGGGGCTGAC CTATCCGCG GCGATGCGCT ATGTGATTCT
501 GCCGCAGGCA TTGCGCCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
551 CGCTCTTGAA AGACAGCTCG CTGCTGTCCG TCATTGCTGT GCGGGAGTTG
601 GCGTATGTT AGAATACGAT TACGGGCCGG TATTCGGTTT ATGAAGAACC
651 GCTTTACACC GTCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
701 GGATATTCTT GCGTTTGAA AACGTTACA ATCCGCAACA CCGCTGA

```

This encodes a protein having amino acid sequence <SEQ ID 840>:

15
20
25
30
35
40
45
50
55
60
65
70

```

1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51 AGAPMRVLAW ALRKVSLLYV TFRGTPLFV QIVIWAYVWF PFFVHPSDGI
101 LVSGEAAIAL RRGYGLIAG SLALIANSQA YICEIFRAGI QSIDKGQMEA
151 ARSLGLTYPQ AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL
201 AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

```

20 ORF129a and ORF129-1 show 100.0% identity in 248 aa overlap:

25
30
35
40

```

orf129a.pep MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
orf129-1 MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW

orf129a.pep ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGLIAG
orf129-1 ALRKVSLLYVTLFRGTPLFVQIVIWAYVWF PFFVHPSDGI LVSGEAAIALRRGYGLIAG

orf129a.pep SLALIANSQA YICEIFRAGI QSIDKGQMEA ARSLGLTYPQ AMRYVILPQA LRRMLPLAS
orf129-1 SLALIANSQA YICEIFRAGI QSIDKGQMEA ARSLGLTYPQ AMRYVILPQA LRRMLPLAS

orf129a.pep EFITLLKDSS LLSVIAVAEL AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE
orf129-1 EFITLLKDSS LLSVIAVAEL AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE

orf129a.pep KRYNPQHRX
orf129-1 KRYNPQHRX

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF129 shows 98.9% identity over a 88 aa overlap with a predicted ORF (ORF129ng) from *N.gonorrhoeae*:

45
50

```

orf129.pep I I Y E Y R W M F L Y G A L T T L G L T V V A X A G G S V L G L L L A L A R L I H L E K A G A P M R V L A W 54
orf129ng M D F R F D I I Y E Y R W M F L Y G A L T T L G L T V V A T A G G S V L G L L L A L A R L I H L E K A G A P M R V L A W 60

orf129.pep A L R K V S L L Y V T L F R G T P L F V Q I V I W A Y V W F P F F V 88
orf129ng A L R K V S L L Y V T L F R G T P L F V Q I V I W A Y V W F P F F V I L H T A F L G N A M R Q S R R V P D K G R W I A G 120

```

An ORF129ng nucleotide sequence <SEQ ID 841> was predicted to encode a protein having amino acid sequence <SEQ ID 842>:

55
60
65
70

```

1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
51 AGAPMRVLAW ALRKVSLLYV TFRGTPLFV QIVIWAYVWF PFFVILHTAF
101 LGNAMQRSRR VPDKGRWIAG SLELNCQPRG RKTRGEFFPG ESNLGTPEPN
151 PLSMGQRRFP GCENWYPPQN FIKK*

```

Further work revealed the following gonococcal sequence <SEQ ID 843>:

```

1 ATGGATTTTc gTTTtGACAT TATTTAcgaA TACCCTGGA TGTTTCTTTA

```

51 CCGGCGCACTG Acgaccttgg ggctgacggt cgtggcgacg gCGGGCGGTT
 101 CGGtattggg TCTGTTGTTG GCGTTGGCGC GCCTGATTCA CTGGGAAAAA
 151 GCGGGTGC GCATGCGCGT GCTGGCGTGG GCGTTGCGTA AGGTTTCGCT
 201 GCTGTACGTT ACCCTGTTCC GGGGTACGCC GCTGTTGTG CAGATTGTGA
 251 TTTGGGCGTA TGTGTGTTT CCGTTTTTCG TCCATCCTTC AGACGGCATT
 301 TTGGTCAGCG GCGAGGCGGC AATCGCGCTG CGTCGCGGAT ACGGGCCGCT
 351 GATTGCCGGT TCTTTGGCAC TGATCGCCAA CTCGGGGGCG TATATCTGTG
 401 AGATTTTCCG CGCGGGCATC CAGTCTATAG ACAAAGGACA GATGGAGGCG
 451 GCGTGTCTT TGGGACTGAC CTATCCGCAG GCGATGCGCT ATGTGATTCT
 501 GCCGAGGCA TTGCGCCGTA TGCTGCCGCC TTTGGCGAGC GAGTTCATCA
 551 CGCTCTTGAA AGACAGCTCG CTGCTGTCGG TCATTGCTGT GCGGGAGTTG
 601 GCGTATGTT AGAATACGAT TACGGGCCGG TATTCGGTT ATGAAGAACC
 651 GCTTTACACC GCCGCCCTGA TTTATCTGTT GATGACGACT TTCTTAGGCT
 701 GGATATTCCT GCGTTTGAA AACGTTACA ATCCGCAACA CCGCTGA

15 This corresponds to the amino acid sequence <SEQ ID 844; ORF129ng-1>:

1 MDFRFDIIYE YRWMFLYGAL TTLGLTVVAT AGGSVLGLLL ALARLIHLEK
 51 AGAPMRVLAW ALRKVSLLYV TFRGTPLFV QIVWAYVWF PFFVHPSDGI
 101 LVSGEAAIAL RRGYGPLIAG SLALIANSYA YICEIFRAGI QSIDKGQMEA
 151 ARSLGLTYPO AMRYVILPQA LRRMLPLAS EFITLLKDSS LLSVIAVAEL
 201 AYVQNTITGR YSVYEEPLYT VALIYLLMTT FLGWIFLRLE KRYNPQHR*

ORF129ng-1 and ORF129-1 show 99.2% identity in 248 aa overlap:

orf129-1.pep MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
 orf129ng-1 MDFRFDIIYEYRWMFLYGALTTLGLTVVATAGGSVLGLLLALARLIHLEKAGAPMRVLAW
 orf129-1.pep ALRKVSLLYVTFRGTPLFVQIVWAYVWFPFFVHPSDGILVSGEAAIALRRGYGPLIAG
 orf129ng-1 ALRKVSLLYVTFRGTPLFVQIVWAYVWFPFFVHPSDGILVSGEAAIALRRGYGPLIAG
 orf129-1.pep SLALIANSYAYICEIFRAGIQSIDKGQMEAAARSLGLTYPOAMRYVILPQALRRMLPLAS
 orf129ng-1 SLALIANSYAYICEIFRAGIQSIDKGQMEAAARSLGLTYPOAMRYVILPQALRRMLPLAS
 orf129-1.pep EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE
 orf129ng-1 EFITLLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLE
 orf129-1.pep KRYNPQHRX
 orf129ng-1 KRYNPQHRX

In addition, ORF129ng-1 is homologous to an ABC transporter from *A. fulgidus*:

2650409(AE001090) glutamine ABC transporter, permease protein (glnP)
 [Archaeoglobus fulgidus] Length = 224
 Score = 132 bits (329), Expect = 2e-30
 Identities = 86/178 (48%), Positives = 103/178 (57%), Gaps = 18/178 (10%)
 Query: 65 VSLLYVTFRGTPLFVQIVWAYVWFPFFVHPSDGILVSGEAAIALRRGYGPLIAGSLAL 124
 +S YV + RGTP VQI+I +F P+ GI + E A G +AL
 Sbjct: 58 ISTAYVEVIRGTPLLVQILI-----VYFGLPAIGINLQPEPA-----GIIAL 99
 Query: 125 IANSYAYICEIFRAGIQSIDKGQMEAAARSLGLTYPOAMRYVILPQALRRMLPLASEFIT 184
 SGAYI EI RAGI+SI GQMEAA SLG+TY QAMRYVI PQA R +LP L +EFI
 Sbjct: 100 SICSGAYIAEIVRAGIESIPIGQMEAAARSLGMYLQAMRYVIFPQAFRNILPALGNEFIA 159
 Query: 185 LLKDSSLLSVIAVAELAYVQNTITGRYSVYEEPLYTVALIYLLMTTFLGWIFLRLEKR 242
 LLKDSSLLSVI++ EL V I P AL YL+MT L + +K+
 Sbjct: 160 LLKDSSLLSVISIVELTRVGRQIVNTTFNAWTPFLGVALFYLMMTIPLSRLVAYSQKK 217

This analysis, including the identification of transmembrane domains in the two proteins, suggests that the proteins from *N. meningitidis* and *N. gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 100

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 845>:

```

1  ..CTGAAAGAAT GCCGTCTGAA AGACCGTGT TTTATTCCAA ATATCGTTTA
51  TAAGAACATC GCCATTACTT TCCTGCTCTT GCACGCCGCC GCCGAACCTT
5  101 GGCTGCCCGC GCAAACCGCC GGTTTTACCG CGCTCGCCGT CGGCTTCATC
151 CTGCTCGCCA AGCTGCGTGA gCTTCACCAT CACGAACCTCT TACGTAAACA
201 cTACGTCCGC ACTTATTACy TGCTCCAAC TTTGCCGCC GCAGgcTAgT
251 TTGTGGACAG GCGCGGCGWA ATTACAAAAC CTGCCCGCyT CCGCGCCCT
301 GCACCTGATT ACCCTCGGCG GCATGATGGG CGGCGTGATG ATGGTGTGGc
10 351 TGACCGCCGG ACTGTGGCAC AGCGGCTTTA CCAAACCTGA CTACCCCAAA
401 CTCTGCCGCA TTGCCGTCCC CATCCTTTTC GCCGCCGCCG TCTCGGCGCG
451 TTTCTTGTG AACGTGAACC CGTATTTT CATTACCGTT CCGCGGATTC
501 TGACCGCCGC CGTATTCGTA CTGTATCTTT TCxCGTTAT ACCGATATTT
551 CGGGCGAATG CGTTTACAGA CGATCCGGAi TAi

```

15 This corresponds to the amino acid sequence <SEQ ID 846; ORF130>:

```

1  ..LKECRKDPV FIPNIVYKNI AITFLLLHAA AELWLPAQTA GFTALAVGFI
51  LLAKLRELHH HELLRKHYVR TYLLQLFAA AGSLWTGAAX LQNLPAAPL
101 HLITLGGMMG GVMVWLTA LWSGFTKLD YPKLCRIAPV ILFAAVSRA
151 FLXNVNPF F ITVPAILTAA VFVLYLFXFI PIFRANAFTD DPE*

```

20 Further work revealed the complete nucleotide sequence <SEQ ID 847>:

```

1  ATGCGGCCGT TTTTCGTGCG CGCGCGGGTG CTTGCCATAC TCGGTGCGCT
51  GGTGTTTTTC ATCAACCCCG GTGCCATCGT CCTGCACCGC CAAATTTTCT
101 TGGAACTTAT GCTGCCGCGC GCATACGGCG GTTTTTTGAC TCGCGCTTGT
151 TTGGACTGGA CGGGTTTTTC GGGTAACCTG AAACCTGTCT CGACTTTGAT
25 201 GCGCGCATT TGTCTGCGCG CATCCGCTAT ACTGCCCTTT TCGCGCAAAA
251 CTGCCTCGTT TTTCTGCGCC GCCTATTGGC TGGTGTGCT GCTGTTCTGC
301 GCGCGGTGA TTTGGCTAGA CCGAAACACC GACAACTTCG CCCTGCTAAT
351 GTTACTTGCC GCGTTCCTG TTTTCAGAC GGCATATGCC GTACGCGCG
401 ATTTGAACCT GTTGCGCGCG CAAGTGATC TAAATATGGC GCGCGTGATG
30 451 TTCGTATCCG TGCGCGTCAG TATTCTTTG GCGCGGAAG CCCTGAAAGA
501 ATGCCGTCTG AAAGACCCGT TTTTATTCC AAATATCGTT TATAAAAACA
551 TCGCCATTAC TTTCTGCTC TTGCACGCGC CCGCGAACT TTGGCTGCC
601 GCGCAAACCG CCGTTTTTAC CGCGCTCGCC GTCGGCTTCA TCCTGCTCGC
651 CAAGCTGCGT GAGCTTCACC ATCAGAACT CTTACGTAAA CACTACGTCC
35 701 GCACTTATTA CCTGCTCCAA CTCTTGCCG CCGCAGGCTA TTTGTGGACA
751 GCGCGGCGGA AATTACAAA CCTGCCCGCC TCCGCGCCCC TGCACCTGAT
801 TACCCTCGGC GGCATGATGG GCGCGTGAT GATGGTGTGG CTGACCGCCG
851 GACTGTGGCA CAGCGGCTTT ACCAACTCG ACTACCCCAA ACTCTGCCG
901 ATTGCCGTCC CCATCCTTTT CGCCGCGCC GTCTCGCGG CTTTCTTGAT
40 951 GAACGTGAAC CCGATATTTT TCATTACCGT TCCTGCGATT CTGACCGCCG
1001 CCGTATTCGT ACTGTATCTT TTCACGTTA TACCGATATT TCGGGCGAAT
1051 GCGTTTACAG ACGATCCGGA ATAA

```

This corresponds to the amino acid sequence <SEQ ID 848; ORF130-1>:

```

1  MRPFFVGA AV LAILGALVFF INPGAIVLHR QIFLEMLLPA AYGGFLTAAL
45 51 LDWTGFSGNL KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVLVLLFC
101 ARLIWLDRNT DNFAILLMLLA AFTVFOTAYA VSGDLNLLRA QVHLNMAAVM
151 FVSVRVSILL GAEALKECRL KDPVFIPNIV YKNIAITFLL LHAAAEWLWP
201 AQTAGFTALA VGFILLAKLR ELHHHELLRK HYVRTYYLLQ LFAAAGYLWT
251 GAAKLQNLPA SAPHLITLG GMMGGVMMVW LTAGLWHS GF TKLDYPKLCR
50 301 IAVPILEAAA VSRAFLMNVN PIFFITVPAI LTAAVFVLYL FTFIPIFRAN
351 AFTDDPE*

```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF130 shows 94.3% identity over a 193aa overlap with an ORF (ORF130a) from strain A of *N.*

55 *meningitidis*:

-463-

```

                    10      20      30
orf130.pep          LKECRLKDPVFIPNIVYKNIAITFLLHAA
                    |||||
orf130a             LNLRAQVHLNMAAVMFVSVRSILLGAEALKECRLKDPVFIPNVYKNIAITFLLHAA
                    140      150      160      170      180      190

                    40      50      60      70      80      90
orf130.pep          AELWLPAQTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGSLWTGAAX
                    |||||
orf130a             AELWLPAQTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYLLQLFAAAGYLWTGAAX
                    200      210      220      230      240      250

                    100      110      120      130      140      150
orf130.pep          LQNLPAAPLHLITLGGMMGVMVWLTAGLWHSFGFTKLDYPKLCRIAPILFAAAVSRA
                    |||||
orf130a             LQNLPAAPLHLITLGGMMGVMVWLTAGLWHSFGFTKLDYPKLCRIAPILFAAAVSRA
                    260      270      280      290      300      310

                    160      170      180      190
orf130.pep          FLXNVNPXFFITVPAILTAAVFVLYLFXFIPIFRANAFTDDPEX
                    |||||
orf130a             VLMNVNPIFFITVPAILTAAVFVLYLLTFVPIFRANAFTDDPEX
                    320      330      340      350

```

The complete length ORF130a nucleotide sequence <SEQ ID 849> is:

```

25      1  ATGCGGCGCT TTTTCGTCGG CGCGGCGGTG CTTGCCATAC TCGGTGCGCT
      51  GGTGTTTTTC ATCAACCCCG GTGCCATCGT CCTGCACCGC CAAATTTTCT
     101  TGGAACTTAT GCTGCCGCGC GCATACGCGC GTTTTTTGAC TCGGGCTTTG
     151  TTGGACTGGA CGGGTTTTTC GGGTAACCTG AAACCTGTCG CGACTTTGAT
     201  GCGGCGATTA TTGCTCGCCG CATCCGCTAT ACTGCCCTTT TCGCCGCAAA
     251  CTGCCTCGTT TTTGCTCGCC GCCTATTGGC TGGTGTGCTG GCTGTTCTGC
     301  GCCCGGCTGA TTTGCTAGA CCGAAACACC GACAACTTCG CCCTGCTAAT
     351  GTTACTTGCC GCGTTCACG TTTTTCAGAC GGCATATGCC GTCAGCGCGC
     401  ATTTGAACCT GTTGC GCGCGC CAAGTGCATC TAAATATGGC GCGGTGATG
     451  TTCGTATCCG TCGCGTCAG TATTCTTTTG GCGCGGAAG CCCTGAAAGA
     501  ATGCGCTCTG AAAGACCCAG TATTCATCCC CAATGTCGTC TATAAAAACA
     551  TCGCCATTAC CTTCTGCTC CTGCACGCGC CCGCCGAAC TTTGGCTGCT
     601  GCGCAACCG CCGGTTTTAC CTCGCTCGCC GTCGGCTTTA TCCTGCTTGC
     651  CAAGCTGCGT GAGCTTCACC ATCAGCAACT CCTGCGCAAA CACTACGTCC
     701  GCACTTATTA CCTGCTCCAA CTCTTTGCCG CCGCAGGCTA TTTGTGGACA
     751  GCGCGGCGCA AATTACAAA CCTGCCGCGC TCCGCGCCCC TGCACCTGAT
     801  TACCCTCGGT GGCATGATGG GCAGCGTGAT GATGGTGTGG CTGACTGCCG
     851  GACTGTGGCA CAGCGGCTTT ACCAAGCTCG ACTACCGGAA ACTCTGCCGC
     901  ATCGCGCTCC CCATCTNTT CGCCGCCGCC GTTTCGCGCG CTGTTTAAAT
     951  GAACGTAAAC CCGATATTCT TCATCACCGT CCGCGCAATT CTGACCGCGC
    1001 CCGTGTTCTG GCTTTACCTG CTGACATTCT TACCGATCTT TCGGGCGAAC
    1051 GCGTTACAG ACGATCCGGA ATAA

```

This encodes a protein having amino acid sequence <SEQ ID 850>:

```

      1  MRPFFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL
     51  LDWTFGSGNL KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVLVLLFC
    101  ARLIWLDRNT DNFAALLMLA AFTVFQAYAY VSGDLNLLRA QVHLNMAAVM
    151  FVSVRSILL GAEALKECRL KDPVFIPNVV YKNIAITELL LHA AELWLP
    201  AQTAGFTSLA VGFILLAKLR ELHHHELLRK HYVRTYLLQ LFAAAGYLWT
    251  GAAKLQNLPA SAPLHLITLG GMMGSVMVW LTAGLWHSGF TKLDYPKLCR
    301  IAVPILEAAA VSRVAVLMNVN PIFFITVPAI LTA AVFVLYL LTFVPIFRAN
    351  AFTDDPE*

```

ORF130a and ORF130-1 show 98.3% identity in 357 aa overlap:

```

      orf130a.pep  MRPFFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL LDWTFGSGNL
      orf130-1    MRPFFVGA AV LAILGALVFF INPGAIVLHR QIFLELMLPA AYGGFLTAAL LDWTFGSGNL
    60      orf130a.pep  KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVLVLLFC ARLIWLDRNT DNFAALLMLA
      orf130-1    KPVATLMAAL LLAASAILPF SPQTASFFVA AYWLVLVLLFC ARLIWLDRNT DNFAALLMLA
    65      orf130a.pep  AFTVFQAYAVSGDLNLLRAQVHLNMAAVMFVSVRSILLGAEALKECRLKDPVFIPNVV

```

-464-

```

      |||
orfl30-1  AFTVFQTAYAVSGDLNLLRAQVHLNMAAVMFVSVRVSILLGAEALKECRLKDPVFIPNIV
      |||
5  orfl30a.pep  YKNIAITFLLHAAELWLPAGTAGFTSLAVGFILLAKLRELHHHELLRKHYVRTYYLLQ
      |||
orfl30-1  YKNIAITFLLHAAELWLPAGTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQ
      |||
10 orfl30a.pep  LFAAAGYLWTGAAKLQNLPAAPLHLITLGGMMGSMVMVWLTAGLWHSAGFTKLDYPKLCR
      |||
orfl30-1  LFAAAGYLWTGAAKLQNLPAAPLHLITLGGMMGSMVMVWLTAGLWHSAGFTKLDYPKLCR
      |||
orfl30a.pep  IAVPILFAAAVSRVIMNVNPIFFITVPAILTAAVFVLYLLTFVPIFRANAFTDDPE
      |||
15 orfl30-1  IAVPILFAAAVSRVIMNVNPIFFITVPAILTAAVFVLYLFTFPIFRANAFTDDPE

```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF130 shows 91.7% identity over a 193 aa overlap with a predicted ORF (ORF130ng) from *N.gonorrhoeae*:

```

20 orfl30.pep                      LKECRLKDPVFIPNIVYKNIAITFLLHAA 30
orfl30ng  LNLRAQVHLNMAAVMFVSVRVSVLLGTETLKECRLKDPVFIPNIVYKNIAIT-LLLHAA 201
orfl30.pep  AELWLPAGTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGSLWTGAAX 90
25 orfl30ng  AELWLPAGTAGFTALAVGFILLAKLRELHHHELLRKHYVRTYYLLQLFAAAGYLWTGAAX 261
orfl30.pep  LQNLPAAPLHLITLGGMMGSMVMVWLTAGLWHSAGFTKLDYPKLCRIAVPILFAAAVSRA 150
orfl30ng  LQNLPAAPLHLITLGGMTGGVMVWLTAGLWHSAGFTKLDYPKLCRIAVSILFASAVSRA 321
30 orfl30.pep  FLXNVNPFITVPAILTAAVFVLYLFXFPIFRANAFTDDPE 193
orfl30ng  VLMNVNPIFFITVPEILTAAVFVLYLFTFVPIFRANAFTDDPE 364

```

An ORF130ng nucleotide sequence <SEQ ID 851> was predicted to encode a protein having amino acid sequence <SEQ ID 852>:

```

1  MNKFFTHPMR PFFVGA AVLA ILGALVFFHQ PRRYHPAPPN FLGTYAAGCI
51 RRFDYRFVFG PDGFFRQPET CRYFDGGVVA CCGCFIAVET ATCRIFRRRL
101 LAGVA AVLRL ADLARRQHRT LRSVDVTA AF TVFQTAYAVS GDLNLLRAQV
151 HLNMAAVMFV SVRVSVLLGT ETLKECRLKD PVFIPNVIYK NIAITLLHHA
40 201 AAELWLPAGT AGFTALAVGF ILLAKLRELH HHELLRKHYV RTYLLQLFA
251 AAGYLWTGAA KLQNLPAAP LHLITLGGMT GGVMVWLTG GLWHSAGFTKL
301 DYPKLCRIAV SILFASAVSR AVLNVNPIF FITVPEILTA AVFVLYLLTF
351 VPIFRANAFT DDPE*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 853>:

```

45 1  ATGCGCCCGT TTTTCGTCGG TCGGCAGTA CTTGCCATAC TCGGTGCGTT
51  GGTGTTTTTT ATCAACCCCG GCGCTATCAT CCTGCACCGC CAAATTTTCT
101 TGGAACTTAT GCTGCCGGCT GCATACGGCG GTTTTGTGAC TACCGCTTGT
151 TTGGACCGGA CGGGTTTTTC AGGCAACCTG AAACCTGCCG CTACTTTGAT
201 GCGCGTGTTG TTGCTGTGTT CGGCTGTTTT ATTGCCGTTT TTACCGCAAC
50 251 TTGCCGCATT TTTCTGCGCC GCCTATTGGC TGGTGTGCT GCTGTCTGTC
301 GCCTGGCTGA TTTGGCTCGA CCGCAACACC GACAACTTCG CTCTGTTGAT
351 GTTACTTGCC GCATTACCG TTTTTCAGAC GGCCTATGCC GTCAGCGCGC
401 ATTTGAACTT ACTGCGCGCG CAAGTGCATT TGAATATGGC GCGGTCATG
451 TTCGTATCCG TCCGCGTCAG CGTCCTTTTG GGCACGAAA CCCTGAAAGA
55 501 ATGCCGCTG AAAGACCCCG TATTCATCCC CAACGTTATC TATAAAAACA
551 TCGCCATCAC CTTGCTGCTG CACGCCGCGG CCGAACTTTG GCTGCCCGCG
601 CAAACCGCCG GTTTTACTGC GCTTGCCGTC GGCTTCATCC TGCTCGCCAA
651 GCTGCGCGAA CTGCACCATC ACGAACTCTT ACGCAAACAC TACGTCCGCA
701 CTTATTACCT GCTCCAGCTC TTTGCCGCGG CAGGTTATCT GTGGACAGGC
60 751 GCGGCGAAAC TGCAAAACCT GCCCGCTCC GCGCCCTGC ACCTGATTAC
801 CCTCGGCGGC ATGACGGGTG GCGTGATGAT GGTGTGGCTG ACTGCCGGAC
851 TGTGGCACAG CGGCTTTACC AAACCTGACT ACCCGAAACT CTGCCGCATC

```

-465-

901 GCCGTCTCCA TCCTTTTCGC CTCCGCCGTT TCGCGCGCTG TTTTAATGAA
 951 CGTGAATCCG ATATTCTTCA TCACCGTTCC CGAGATTCTG ACCGCCGCCG
 1001 TGTTTCATGCT TTACCTGCTG ACGTTCGTAC CGATTTTTCG AGCGAACGCG
 1051 TTTACAGACG ATCCGAATA A

5 This corresponds to the amino acid sequence <SEQ ID 854; ORF130ng-1>:

1 MRPFVGA AV LAILGALVFF INPGA IILHR QIFLELM LPA AYGGFLT TAL
 51 LDRTGFSGNL KPAATLMAVL LLVA AVLLPF LPQLAAFFVA AYWLVL LLLFC
 101 AWLIWLD RNT DNFA LLMLLA AFTVFQ TAYA VSGDLN LLRA QVHLNMAA VM
 151 FVSVRVS VLL GTETLKEC RL KDPVFIP NVI YKNIAIT LLL HAAAE L W LPA
 201 QTAGFTAL AV GFILLAKL RE LHHHELLR KH YVRTYLL QL FAAAGYL WTG
 251 AAKLQNL PAS APLHLITL GG MTGGVM MVWL TAGLWHS GFT KLDYPKL CRI
 301 AVSILFAS AV SRAVL MNVNP IFFITV PEIL TAAVFM LYL T FVPIFRANA
 351 FTDDPE*

ORF130ng-1 and ORF130-1 show 92.4% identity in 357 aa overlap:

15 orf130-1.pep MRPFVGA AV LAILGALVFF INPGA IIVLHRQIFLELM LPAAYGGFLTA ALLDWTGFSGNL
 orf130ng-1 MRPFVGA AV LAILGALVFF INPGA IILHRQIFLELM LPAAYGGFLTALLDRTGFSGNL
 20 orf130-1.pep KPVATLMA ALLLAASAILPFSPQTASFFVAAYWLVL LFCARLIWLD RNTDNFA LLMLLA
 orf130ng-1 KPAATLMA VLLLVAAVLLPFLPQLAAFFVAAYWLVL LFCAWLIWLD RNTDNFA LLMLLA
 25 orf130-1.pep AFTVFQ TAYAVSGDLN LLRAQVHLNMAA VMFVSVRVS ILLGAEALKEC RLKDPVFIP NVI
 orf130ng-1 AFTVFQ TAYAVSGDLN LLRAQVHLNMAA VMFVSVRVS VLLGTETLKEC RLKDPVFIP NVI
 30 orf130-1.pep YKNIAIT FLLHAAAE L W LPAQTAGFTAL AVGFILLAKL RELHHHELLR KH YVRTYLL Q
 orf130ng-1 YKNIAIT -LLLHAAAE L W LPAQTAGFTAL AVGFILLAKL RELHHHELLR KH YVRTYLL Q
 35 orf130-1.pep LFAAAGYL WTGA AKLQNL PASAPLHLITLGGMMGVM MVWL TAGLWHS GFTKLDYPKL CR
 orf130ng-1 LFAAAGYL WTGA AKLQNL PASAPLHLITLGGMTGGVM MVWL TAGLWHS GFTKLDYPKL CR
 40 orf130-1.pep IAVPILFAAAVSRAFL MNVNP IFFITVPA ILTA AVFVLYLFTFIPIFRANAFTDDPEX
 orf130ng-1 IAVSILFAS AVSRAVL MNVNP IFFITVPEILTA AVFMYLLT FVPIFRANAFTDDPEX

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

40 Example 101

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 855>:

1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCA TT
 51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA
 101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATT TTTGGGATATT
 151 GCGGCGGAGA GTCCGCCGTC TTAGGGGAC TACGAGATAC CGCTTTT CAGA
 201 CGGCAATAGT TCCGT CAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
 251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG C.TGCGGGCT GGATTGGCGT
 301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTTGA
 351 CTGCTTGGA AAG..

50 This corresponds to the amino acid sequence <SEQ ID 856; ORF131>:

1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI
 51 GGESPPSLGD YEIPLSDGNS SVRANEYESA QQSYFYRKIG KFEXCGLDWR
 101 TRDGKPLIET FKQGGFDCLE K..

Further work revealed the complete nucleotide sequence <SEQ ID 857>:

55 1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCA TT
 51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCCTCA

-466-

5
101 CCGGCTGGTG TAAGCCGAGA AAACCGGCTG CCATCGATT TTGGGATATT
151 GCGGCGGAGA GTCCGCCGTC TTTAGGGGAC TACGAGATAC CGCTTTCAGA
201 CCGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGCT GGATTGGCGT
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG GAGGATTGTA
351 CTGCTTGAA AAGCAGGGGT TCGGCGCAA CGGTCTGTCC GAGCGCGTCC
401 GATGGTAA

This corresponds to the amino acid sequence <SEQ ID 858; ORF131-1>:

10
1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLTGWCKPR KPAAIDFWDI
51 GGESPPSLGD YEIPLSDGNNR SVRANEYESA QQSYFYRKIG KFEACGLDWR
101 TRDGKPLIET FKQGGFDCLE KQGLRRNGLS ERVRW*

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF131 shows 95.0% identity over a 121aa overlap with an ORF (ORF131a) from strain A of *N.*

15 *meningitidis*:

		10	20	30	40	50	60
orf131.pep	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD						
orf131a	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLED						
20		10	20	30	40	50	60
orf131.pep	YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCLE						
25	orf131a	YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLE					
		70	80	90	100	110	120
30	orf131.pep	K					
	orf131a	KQGLRRNGLSERVRWX					
		130					

The complete length ORF131a nucleotide sequence <SEQ ID 859> is:

35
1 ATGGAAATTC GGGCAATAAA ATATACGGCA ATGGCTGCGT TGCTTGCAAT
51 TACGGTTGCA GGCTGCCGGT TGGCAGGTTG GTATGAGTGT TCGTCCCTGT
101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATT TTGGGATATT
151 GCGGCGGAGA GTCCTCCGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA
201 CCGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCA CAACAATCTT
251 ACTTTTACAG GAAAATAGGG AAGTTTGAAG CCTGCGGGTT GGATTGGCGT
40
301 ACGCGTGACG GCAAACCTTT GATTGAGACG TTCAAACAGG AAGGTTTGA
351 TTGTTTGAAA AAGCAGGGGT TCGGCGCAA CGGTCTGTCC GAGCGCGTCC
401 GATGGTAA

This encodes a protein having amino acid sequence <SEQ ID 860>:

45
1 MEIRAIKYTA MAALLAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI
51 GGESPPSLED YEIPLSDGNNR SVRANEYESA QQSYFYRKIG KFEACGLDWR
101 TRDGKPLIET FKQEGFDCLE KQGLRRNGLS ERVRW*

ORF131a and ORF131-1 show 97.0% identity in 135 aa overlap:

	orf131a.pep	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPPSLED
50	orf131-1	MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD
	orf131a.pep	YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQEGFDCLE
	orf131-1	YEIPLSDGNRSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE
55	orf131a.pep	KQGLRRNGLSERVRWX

orf131-1 KQGLRRNGLSERVRWX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF131 shows 89.3% identity over 121 aa overlap with a predicted ORF (ORF131ng) from

5 *N.gonorrhoeae*:

```

orf131.pep  MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD  60
            |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:
orf131ng    MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPLSLED  60

10 orf131.pep  YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEXCGLDWRTRDGKPLIETFKQGGFDCLE  120
            |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:
orf131ng    YEIPLSDGNSSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE  120

15 orf131.pep  K 121
            |
orf131ng      KQGLRRNGLSERVRW 134

```

A complete length ORF131ng nucleotide sequence <SEQ ID 861> was predicted to encode a protein having amino acid sequence <SEQ ID 862>:

```

1  MEIRVIKYTA TAALFAFTVA GCRLAGWYEC LSLSGWCKPR KPAAIDFWDI
20 51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

```

Further work revealed the following gonococcal DNA sequence <SEQ ID 863>:

```

1  ATGGAATTC GGGTAATAAA ATATACGGCA ACGGCTGCGT TGTTCGATT
25 51 TACGGTTGCA GGCTGCCGGC TGGCGGGGTG GTATGAGTGT TCGTCCTTGT
101 CCGGCTGGTG TAAGCCGAGA AAACCTGCCG CCATCGATT TTGGGATATT
151 GCGGCGCAGA GtcgcgtGTC TTTAGAGGAC TACGAGATAC CGCTTTCAGA
201 CGGCAATCGT TCCGTCAGGG CAAACGAATA TGAATCCGCG CAAAAATCTT
251 ACTTTTATAG GAAAATAGGG AAGTTGAAG CCTGCGGGTT GGATTGGCGT
301 ACGCGTGACG GCAAACCTTT GGTGAGAGG TTCAAACAGG AAGGTTTCGA
30 351 CTGTTTGAA AAGCAGGGT TCGGCGCAA CGCCTGTCC GAGCGCGTCC
401 GATGGTAA

```

This corresponds to the amino acid sequence <SEQ ID 864; ORF131ng-1>:

```

1  MEIRVIKYTA TAALFAFTVA GCRLAGWYEC SSLSGWCKPR KPAAIDFWDI
35 51 GGESPLSLED YEIPLSDGNR SVRANEYESA QKSYFYRKIG KFEACGLDWR
101 TRDGKPLVER FKQEGFDCLE KQGLRRNGLS ERVRW*

```

ORF131ng-1 and ORF131-1 show 92.6% identity in 135 aa overlap:

```

orf131ng-1.pep MEIRVIKYTATAALFAFTVAGCRLAGWYECSSLSGWCKPRKPAAIDFWDIGGESPLSLED
40 orf131-1      MEIRAIKYTAMAALLAFTVAGCRLAGWYECSSLTGWCKPRKPAAIDFWDIGGESPPSLGD

orf131ng-1.pep YEIPLSDGNSSVRANEYESAQKSYFYRKIGKFEACGLDWRTRDGKPLVERFKQEGFDCLE
orf131-1        YEIPLSDGNSSVRANEYESAQQSYFYRKIGKFEACGLDWRTRDGKPLIETFKQGGFDCLE

45 orf131ng-1.pep KQGLRRNGLSERVRWX
            |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:||||| |||:
orf131-1        KQGLRRNGLSERVRWX

```

Based on the presence of a predicted prokaryotic membrane lipoprotein lipid attachment site, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be

50 useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 102

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 865>

```

1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
51  TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
151 TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
251 TGAACCTCGG CCTGCCtTAT ATtTcCGGCC CGCAATGGCT GTCGGAAAAC
301 GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACgC ACGGCAAAAC
10 351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATgCC GGCCTCGCGC
401 CGGGCTTCCT TATtGGCGGC GTACC.GGAA AATtTcGGCG TTTCCGCCCG
451 CCTGCCGCAA ACGCCGCGCC AAGACCGGAA CAGCCAATCG CCGTTTTTcG
501 TCATCGAAGC CGACGAATAC GACACCGCCT TTtTCGACAA ACGTTCTAAA
551 TtCGTGCATT ACCGTCGCGG TACCGCCGTG TTGAACAATC TGAATTTCGA
15 601 CCACGCCGAC ATCTTTGCCG ACTTGGGCGC GATACAGAcC CAGTTCCTACT
651 ACCTCGTGCG TACCGTGCCG TCTGAAGGCT TAATCGTCTG CAACGGACGG
701 CAGCAAAGCC TGCAAGATAC TTTGGACAAA GGCTGCTGGA CGCCGGTGGa
751 AAAATTCGGC ACGGAACACG GCTGGCA..

```

This corresponds to the amino acid sequence <SEQ ID 866; ORF132>:

```

20 1  MKHIHIIGIG GTFMGGAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
51  YEGFDAAQLD EFKADVYVIG NVAKRGMVV EAILNLGLPY ISGPQWLSEN
101 VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VXGKFRFRFP
151 PAANAAPRFE QPIAVFRHRS RRIHRHLFRQ TFXIRALPSA YRRVEQSGIR
201 PRRHLCRLGR DTDVPLPRA YRAVXRLNRL QRTAAKPARY FGQRLLDAGG
25 251 KIRHGTRLA..

```

Further work revealed the complete nucleotide sequence <SEQ ID 867>:

```

1  ATGAAACACA TCCATATTAT CGGTATCGGC GGCACGTTTA TGGGCGGGCT
51  TGCCGCCATT GCCAAAGAAG CGGGGTTTGA AGTCAGCGGT TGCGACGCGA
101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG TATAGACGTG
30 151 TATGAAGGCT TCGATGCCGC TCAGTTGGAC GAATTTAAAG CCGACGTTTA
201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATTT
251 TGAACCTCGG CCTGCCtTAT ATtTcCGGCC CGCAATGGCT GTCGGAAAAC
301 GTGCTGCACC ATCATTGGGT ACTCGGTGTG GCGGGGACgC ACGGCAAAAC
35 351 GACCACCGCC TCCATGCTCG CATGGGTCTT GGAATATGCC GGCCTCGCGC
401 CGGGCTTCCT TATtGGCGGC GTACCGGAAA ATtTcGGCGT TtCCGCCCGC
451 CTGCCGCAAA CGCCGCGCCA AGACCGGAAC AGCCAATCGC CCGTTTTTCG
501 CATCGAAGCC GACGAATACG ACACCGCCTT TTTCGACAAA CGTTCTAAAT
551 TCGTGCATTA CCGTCGCGGT ACCGCCGTGT TGAACAATCT GGAATTCGAC
601 CACGCCGACA TCTTTGCCGA CTTGGGCGCG ATACAGACCC AGTTCCTACT
40 651 CCTCGTGCGT ACCGTGCCGT CTGAAGGCTT AATCGTCTGC AACGGACGGC
701 AGCAAAGCCT GCAAGATACT TTGGACAAAG GCTGCTGGAC GCCGGTGGAA
751 AAATTCGGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCGACGG
801 CTCGTTGACG GTGTTGCTCG ACGGCAAAAC CGCCGGACGC GTCAAAATGGG
851 ATTTGATGGG CAGGCACAAC CGCATGAACG CGCTCGCGGT CATTGCGGCC
45 901 GCGCGTCATG TCGGTGTGCA TATTCAGACC GCCTGCGAAG CCTTGGGCGC
951 GTTTAAAAAC GTCAAACGCC GGATGGAAT CAAAGGCACG GCAAACGGCA
1001 TCACCGTTTA CGACGACTTC GCCCACCACC CGACCGCCAT CGAAACCACG
1051 ATTCAGGTTT TGCGCCAACG CGTCGGCGGC GCGCGCATCC TCGCCGTCCT
1101 CGAACC CGT TCCAACACGA TGAAGCTGGG CACGATGAAG TCCGCCCTGC
50 1151 CTGTAAGCCT CAAAGAAGCC GACCAAGTGT TCTGCTACGC CGCGCGCGTG
1201 GACTGGGACG TCGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGAACGT
1251 CGGCAAAGAC TTCGATGCCT TCGTTGCCGA AATCGTGAAG AACGCCGAAG
1301 TAGGCGACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
1351 GGAAAGCTGC TGGAAGCTTT GAGATAG

```

55 This corresponds to the amino acid sequence <SEQ ID 868; ORF132-1>:

```

1  MKHIHIIGIG GTFMGGAAI AKEAGFEVSG CDAKMYPPMS TQLEALGIDV
51  YEGFDAAQLD EFKADVYVIG NVAKRGMVV EAILNLGLPY ISGPQWLSEN
101 VLHHHWVLGV AGTHGKTTTA SMLAWVLEYA GLAPGFLIGG VPENFGVSAR
151 LPQTFRQDPN SQSPFFVIEA DEYDTAFFDK RSKFVHYRPR TAVLNNLEFD
60 201 HADIFADLGA IQTQFHYLVR TVPSEGLIVC NGRQOQLQDT LDKGCWTFVE
251 KFGTEHGWA GEANADGSFD VLLDGKTAGR VKWDLGRHN RMNALAVIAA
301 ARHVGVDIQT ACEALGAFKN VKRMEIKGT ANGITYDDF AHHPATAIET

```

351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK SALPVSLKEA DQVFCYAGGV
 401 DWDVAEALAP LGGRLNVGKD FDAFVAEIVK NAEVGDHILV MSNGGFGGIH
 451 GKLEALR*

Computer analysis of this amino acid sequence gave the following results:

5 Homology with the hypothetical o457 protein of *E. coli* (accession number U14003)

ORF132 and o457 show 58% aa identity in 140 aa overlap:

10 Orf132: 4 IHIIGIGGTFMGGGLAAIAKEAGFEVSGCDKMYPPMSTQLEALGIDVYEGFDAAQLDEFK 63
 IHI+GI GTFMGGGLA +A++ G EV+G DA +YPPMST LE GI++ +G+DA+QL+ +
 o457: 3 IHIIGIGGTFMGGGLAIAKEAGFEVSGCDKMYPPMSTLLEKQIELIQYDASQLEP-Q 61
 Orf132: 64 ADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTASML 123
 D+ +IGN RG VEA+L +PY+SGPQWL + VL WVL VAGTHGKTTTA M
 o457: 62 PDLVIIGNAMTRGNPCVEAVLEKNIPYMSGPQWLHDFVLRDRWVLAAGTHGKTTTAGMA 121
 15 Orf132: 124 AWWLEYAGLAPGFLIGGVXG 143
 W+LE G PGF+IGGV G
 o457: 122 TWILEQCGYKPGFVIGGVPG 141

Homology with a predicted ORF from *N. meningitidis* (strain A)

20 ORF132 shows 74.6% identity over a 189aa overlap with an ORF (ORF132a) from strain A of *N. meningitidis*:

25 orf132.pep MKHHIIGIGGTFMGGGLAAIAKEAGFEVSGCDKMYPPMSTQLEALGIDVYEGFDAAQLD
 orf132a MKHHIIGIGGTFMGGGLAAIAKEAGFEVSGCDKMYPPMSTQLEALGIDVYEGFDAAQLD
 30 orf132.pep EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA
 orf132a EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA
 35 orf132.pep SMLAWVLEYAGLAPGFLIGGVXGKFR---RFRPPAANAAPRPEQPI-----AVFR
 orf132a SMLAWVLEYAGLAPGFXIGGVENFSVSARL-PQTPRODPNSQSPFFVIEADEYDTAFDD
 40 orf132.pep HRSRRIRHRLFRQTFXIRALPSAYRRVEQSGIRPRRHLCRLGRDTPVPLPRAYRAVXRL
 orf132a KRSKFVHYRPTAVLNNLEFDHADIFADLGAIQTQFHHLVVRTVPSEGLIVCNGRQQSLOD

45 The complete length ORF132a nucleotide sequence <SEQ ID 869> is:

1 ATGAAACACA TCCACATTAT CGGTATCGGC GGCACGTTTA TGGGTGGGAT
 51 TGCCGCCATT GCCAAGAAG CAGGGTTTGA ANTCAGCGGT TGCGATGCCA
 101 AGATGTATCC GCCGATGAGC ACCCAGCTCG AAGCCTTGGG CATAGGCGTG
 151 TATGAAGGCT TCGACACCGC GCAGTTGGAC GAATTTAAAG CCGACGTTTA
 201 CGTTATCGGC AATGTCGCCA AGCGCGGGAT GGATGTGGTT GAAGCGATT
 251 TGAACCGTGG GCTGCCTTAT ATTTCCGGCC CGCAATGGCT GGCTGAAAAC
 301 NTGCTGCACC ATCATTGGNN ACTCGGCGTG GCGGNGACGC ACGGCAAAAC
 351 GACCACCGCG TCTATGCTCG CGTGGGTTT GGAATATGCC GGAATCGCAC
 401 CGGGCTTCNT TATCGGCGGC GTACCGGAAA ACTTCAGCGT TTCCGCCCGC
 451 CTGCCGCAAA CGCCGCGCCA AGACCGAAC AGCCAATCGC CGTTTTCGT
 501 CATTGAAGCC GACGAATACG ACACCGCGT TTTGACAAA CGTCCAAAT
 551 TCGTGCATTA CCGTCCGCGT ACCGCGGTGT TGAACAATCT GGAATTCGAC
 601 CACGCCGACA TCTTCGCCGA TTTGGGCGCG ATACAGACCC AGTTCACCA
 651 CCTCGTGGT ACCGTGCCGT CTGAAGGCCT CATCGTCTGC AACGGACGGC
 701 AGCAAAGCCT GCAAGACACT TTGGACAAAG GCTGCTGGAC GCCGCTGGAA
 751 AAATTCGGCA CGGAACACGG CTGGCAGGCC GGCGAAGCCA ATGCCGATGG

-470-

801 CTCGTTTCGAC GTGTTGCTTG ACGGCAAAAA AGCCGGACAC GTCGCTTGGA
 851 GTTTGATGGG CGGACACAAC CGCATGAACG CGCTCGCNGT CATCGCCGCC
 901 GCGCGTCATG CCGGAGTNGA CATTGAGACG GCCTGCGAAG CCTTGAGCAC
 951 GTTTAAAAAC GTCAAAACGCC GCATGGAAAT CAAAGGCACG GCAAACGGTA
 1001 TCACCGTTTA CGACGACTTC GCCCACCATC CGACCGCTAT CGAAACACAG
 1051 ATTCAGGTT TGCGCCAGCG CGTCGGCGGC GCGCGCATCC TCGCCGTCCT
 1101 CGAACC CGCT TCCAATACGA TGAAGCTGGG TACGATGAAA GCCGCCCTGC
 1151 CCGCAAGCCT CAAAGAAGCC GACCAAGTGT TCTGNTACGC CGCGGCGCGC
 1201 GACTGGGACG TTGCCGAAGC CCTCGCGCCT TTGGGCGGCA GGCTGCACGT
 1251 CGGCAAAGAC TTCGATGCCT TCGTTGCCGA AATCGTGAAA AACGCCGAAG
 1301 CAGCGCACCA TATTTTGGTG ATGAGCAACG GCGGTTTCGG CGGAATACAC
 1351 ACCAAACTGC TGGACGCTTT GAGATAG

This encodes a protein having amino acid sequence <SEQ ID 870>:

1 MKHIHIIGIG GTFMGGIAAI AKEAGFEXSG CDAKMYPPMS TQLEALGIGV
 51 YEGFDTAQLD EFKADVYVIG NVAKRGM DVV EAILNRGLPY ISGPQWLAEN
 101 XLHHHWXLGV AXTHGKTTTA SMLAWVLEYA GLAPGFXIGG VPENFVSAR
 151 LPQTPRQDPN SOSFFVIEA DEYDTAFD KRSKFVHYRPR TAVLNNLEFD
 201 HADIFADLGA IQTQFHHLVR TVPSEGLIVC NGRQOSLQDT LDKGCWTPVE
 251 KFGTEHWQAG GEANADGSFD VLLDGKKAGH VAWSLMGHNM RNALAVIAA
 301 ARHAGVDIQT ACEALSTFKN VKRMEIKGT ANGITYDDF AHHPTAIETT
 351 IQGLRQRVGG ARILAVLEPR SNTMKLGTMK AALPASLKEA DQVFXYAGGA
 401 DWDVAEALAP LGGRLHVGKD FADFVAEIVK NAEAGDHILV MSNGGFGGIH
 451 TKLLDALR*

ORF132a and ORF132-1 show 93.9% identity in 458 aa overlap:

25 orf132a.pep MKHIHIIGIGGTFMGGIAAIKEAGFEXSGCDAKMYPPMSTQLEALGIGVYEGFDTAQLD
 orf132-1 MKHIHIIGIGGTFMGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD
 30 orf132a.pep EFKADVYVIGNVAKRGM DVVEAILNRGLPYISGPQWLAENXLHHHWXLGVAXTHGKTTTA
 orf132-1 EFKADVYVIGNVAKRGM DVVEAILNLGLPYISGPQWLSENVLHHHWLVGVAGTHGKTTTA
 35 orf132a.pep SMLAWVLEYAGLAPGFXIGGVPENFVSARLPQTPRQDPNSQSFFVIEADEYDTAFD K
 orf132-1 SMLAWVLEYAGLAPGFLIGGVPENFGVSARLPQTPRQDPNSQSFFVIEADEYDTAFD K
 40 orf132a.pep RSKFVHYRPR TAVLNNLEFDHADIFADLGA IQTQFHHLVRTVPSEGLIVCNGRQOSLQDT
 orf132-1 RSKFVHYRPR TAVLNNLEFDHADIFADLGA IQTQFHYLVRTVPSEGLIVCNGRQOSLQDT
 45 orf132a.pep LDKGCWTPVEKFGTEHWQAGEANADGSFDVLLDGKKAGHVAWSLMGGHNM RNALAVIAA
 orf132-1 LDKGCWTPVEKFGTEHWQAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRNALAVIAA
 50 orf132a.pep ARHAGVDIQTACEALSTFKNVKRMEIKGTANGITYDDFAHHPTAIETT IQGLRQRVGG
 orf132-1 ARHVGVDIQTACEALGAFKNVKRMEIKGTANGITYDDFAHHPTAIETT IQGLRQRVGG
 55 orf132a.pep ARILAVLEPR SNTMKLGTMKAALPASLKEADQVFXYAGGADWDVAEALAPLGGR LHVGKD
 orf132-1 ARILAVLEPR SNTMKLGTMKSALPVSLKEADQVFCYAGGVDWDVAEALAPLGGR LHVGKD
 orf132a.pep FADFVAEIVKNAEAGDHILVMSNGGFGGIHTKLLDALRX
 orf132-1 FADFVAEIVKNAEAGDHILVMSNGGFGGIH GKLLDALRX

Homology with a predicted ORF from *N.gonorrhoeae*

ORF132 shows 89.6% identity over 259 aa overlap with a predicted ORF (ORF132ng) from *N. gonorrhoeae*:

60 orf132.pep MKHIHIIGIGGTFMGGLAAIAKEAGFEVSGCDAKMYPPMSTQLEALGIDVYEGFDAAQLD 60
 orf132ng MKHIHIIGIGGTFMGGIAAIKEAGFKVSGCDAKMYPPMSTQLEALGIGVHEGFDAQA 60

-471-

	orf132.pep	EFKADVYVIGNVAKRGMDVVEAILNLGLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA	120
	orf132ng	EFQADIYVIGNVARRGMDVVEAILNRGLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTA	120
5	orf132.pep	SMLAWVLEYAGLAPGFLIGGVXGKFRFRPPAANAAPRPEQPIAVFRHRSRRIRHRLFRQ	180
	orf132ng	SMLAWVLEYAGLAPGFLIGGVPGKFRFRPPTANAASRPEQQIAVFRHRSRRIRHRLFRQ	180
10	orf132.pep	TFXIRALPSAYRRVEQSGIRPRRHLRRLGRDTPVPLPRAYRAVXRLNRLQRTAAKPARY	240
	orf132ng	TLQIRALSPAYRRVEQSGIRPRRHLRRLGRDTPVPPRAHRTIRRPHRLQRTAAKPARY	240
	orf132.pep	FGQRLLDAGGKIRHGTRLA	259
15	orf132ng	FGQRLLDAGGKIRHRTLADW	261

An ORF132ng nucleotide sequence <SEQ ID 871> was predicted to encode a protein having amino acid sequence <SEQ ID 872>:

	1	MKHIHIIGIG	GTfMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TOLEALGIGV
20	51	HEGFDAQLE	EFQADIYVIG	NVARRGMDVV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWVLGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPGKFRFRP
	151	PTANAASRPE	QQIAVFRHRS	RRIRHRLFRQ	TLQIRALSPA	YRRVEQSGIR
	201	PRRHLRRLGR	DTDPVPPRA	HRTIRRPHRL	QRTAAKPARY	FGQRLLDAGG
	251	KIRHRTLAD	W*			

Further work revealed the following gonococcal DNA sequence <SEQ ID 873>:

25	1	ATGAAACACA	TCCACATTAT	CGGTATCGGC	GGCACGTTTA	TGGGCGGGAT
	51	TGCCGCCATT	GCCAAAGAAG	CCGGGTTCAA	AGTCAGCGGT	TGCGACGCGA
	101	AGATGTATCC	GCCGATGAGC	ACCCAGCTCG	AAGCCTTGGG	CATAGGCGTA
	151	CACGAAGGCT	TCGATGCCGC	GCAAGTTGGAA	GAATTTCAGG	CCGATATTTA
	201	CGTCATCGGC	AATGTCGCCA	GGCGCGGGAT	GGATGTGGTC	GAGGCGATTT
30	251	TGAACCGTGG	GCTGCGCTTAT	ATTTCCGGCC	CGCAATGGCT	GGCTGAAAac
	301	GTGCTgcacc	atcaTTGGgt	ACTCGGCGTG	GcagggaCGC	ACGScaaAac
	351	gaccaCcGcg	tCCATGCTCG	CCTGGGTCTT	GGAATATGCC	GGACTCGCGC
	401	CGGGCTTCCT	CATCGGCGGT	gtaccggaAA	ATTTCCGGCGT	TTCGCGCCGC
	451	CTACCGCAAA	CGCCGCGTCA	AGACCCGAAC	AGCAAATCGC	CGTTTTTCGT
35	501	CATCGAAGCC	GACGAATACG	ACACCGCCTT	TTTCGACAAA	CGCTCCAAAT
	551	TCGTGCATTA	TCGCCCGCGT	ACCGCCGTGT	TGAACAATCT	GGAATTCGAC
	601	CACGCCGACA	TCTTCGCCGA	CTTGGGCGCG	ATACAGACCC	AGTTCACCA
	651	CCTCGTGCGC	ACCGTACCAT	CCGAAGGCCT	CATCGTCTGC	AACGGACAGC
	701	AGCAAAGCCT	GCAAGATACT	TTGGACAAAG	GCTGCTGGAC	GCCGGTGGAA
40	751	AAATTCGGCA	CCGACACGGG	CTGGCAGATT	GGTGAAGTCA	ATGCCGACGG
	801	CTCGTTCGAC	GTATTGCTTG	ACGGCAAAA	AGCCGGACAC	GTGCGATGGG
	851	ATTTGATGGG	CGGACACAAC	CGCATGAACG	CGCTCGCCGT	CATCGCTGCC
	901	GCACGCCATG	CCGGAGTCGA	TGTTCAAGCG	GCCTGCGAAG	CCTTGGGTGC
	951	GTTTAAAAAC	GTCAAACGCC	GCATGGAAT	CAAAGGCACG	GCAAACGGCA
45	1001	TCACCGTTTA	CGACGATTTT	GCCCACCACC	CGACCGCCAT	GAAACACAG
	1051	ATTCAAGGTT	TGCGCCAACG	TGTCGGCGGC	GCGCGCATCC	TCGCGTCTCT
	1101	CGAGCCGCGT	TCCAACACCA	TGAAACTCGG	CACGATGAAG	TCCGCCCTGC
	1151	CCGCAAGCCT	CAAAGAAGCC	GACCAAGTGT	TCTGCTACGC	CGGCGGCGCG
	1201	GACTGGGACG	TTGCCGAAGC	CCTCGCGCCT	TTGGGCTGCA	GGCTGCGCGT
50	1251	CGGTAAAGAT	TTGATACCT	TCGTTGCCGA	AATTGTGAAA	AACGCCCGAA
	1301	CCGGCGACCA	TATTTTGGTG	ATGAGCAACG	GCGGTTTCGG	CGGAATACAC
	1351	ACCAAACGTC	TGGACGCTTT	GAGATAG		

This corresponds to the amino acid sequence <SEQ ID 874; ORF132ng-1>:

	1	MKHIHIIGIG	GTfMGGIAAI	AKEAGFKVSG	CDAKMYPPMS	TOLEALGIGV
55	51	HEGFDAQLE	EFQADIYVIG	NVARRGMDVV	EAILNRGLPY	ISGPQWLAEN
	101	VLHHHWVLGV	AGTHGKTTTA	SMLAWVLEYA	GLAPGFLIGG	VPENFGVSAR
	151	LPQTPRQDPN	SKSPFFVIEA	DEYDTAFFDK	RSKFVHYRPR	TAVLNNLEFD
	201	HADIFADLGA	IQTQFHLVR	TVPSEGLIVC	NGQQSLQDT	LDKGCWTPVE
	251	KFGTGHGWOI	GEVNADGSFD	VLLDGKKAGH	VAWDLMGHNN	RMNALAVIAA
60	301	ARHAGVDVQT	ACEALGAFKN	VKRRMEIKGT	ANGITVYDDF	AHHPTAIETT
	351	IQGLRQRVGG	ARILAVLEPR	SNTMKLGTMK	SALPASLKEA	DQVFCYAGGA
	401	DWDVAEALAP	LGCRRLRVGKD	FDTFVAEIVK	NARTGDHILV	MSNGGFGGIH
	451	TKLLDALR*				

-472-

ORF132ng-1 and ORF132-1 show 93.2% identity in 458 aa overlap:

```

5  orf132ng-1.pep MKHIHIIGIGGTFMGGIAAIAKEAGFKVSGCDKMYPPMSTQLEALGIGVHEGFDAQAQLE
   orf132-1       MKHIHIIGIGGTFMGGIAAIAKEAGFEVSGCDKMYPPMSTQLEALGIDVYEGFDAQQLD

10 orf132ng-1.pep EFQADIYVIGNVARRGMDVVEAILNRLPYISGPQWLAENVLHHHWVLGVAGTHGKTTTA
   orf132-1       EFKADVYVIGNVAKRGMDVVEAILNRLPYISGPQWLSENVLHHHWVLGVAGTHGKTTTA

15 orf132ng-1.pep SMLAWVLEYAGLAPGFLIGVPEPFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDK
   orf132-1       SMLAWVLEYAGLAPGFLIGVPEPFGVSARLPQTPRQDPNSQSPPFFVIEADEYDTAFFDK

20 orf132ng-1.pep RSKFVHYRPRTAVLNNLEFDHADI FADLGAIQTQFHHLVVRTVPSEGLIVCNGQQQSLQDT
   orf132-1       RSKFVHYRPRTAVLNNLEFDHADI FADLGAIQTQFHYLVVRTVPSEGLIVCNGRQQSLQDT

25 orf132ng-1.pep LDKGCWTPVEKFGTGHGWQIGEVNADGSFDVLLDGKKAGHVAWDLMGHNRNMALAVIAA
   orf132-1       LDKGCWTPVEKFGTEHGWQAGEANADGSFDVLLDGKTAGRVKWDLMGRHNRNMALAVIAA

30 orf132ng-1.pep ARHAGVDVQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG
   orf132-1       ARHVGVDIQTACEALGAFKNVKKRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG

35 orf132ng-1.pep ARILAVLEPRSNMTKLGTMKSALPASLKEADQVFCYAGGADWDVAEALAPLGCRLRVGKD
   orf132-1       ARILAVLEPRSNMTKLGTMKSALPVSLKEADQVFCYAGGVDWDVAEALAPLGGRLNVGKD

40 orf132ng-1.pep FDTFVAEIVKNARTGDHILVMSNGGFGGIHTKLLDALRX
   orf132-1       FDAFVAEIVKNAEVDHILVMSNGGFGGIHGKLLLEALRX

```

In addition, ORF132ng-1 is homologous to a hypothetical *E.coli* protein:

```

35  pir||S56459 hypothetical protein o457 - Escherichia coli >gi|537075 (U14003)
   ORF_o457 [Escherichia coli] >gi|1790680 (AE000494) hypothetical 48.5 kD protein
   in fbp-pmba intergenic region [Escherichia coli] Length = 457
   Score = 474 bits (1207), Expect = e-133
   Identities = 249/439 (56%), Positives = 294/439 (66%), Gaps = 13/439 (2%)

40  Query: 22 KEAGFKVSGCDKMYPPMSTQLEALGIGVHEGFDAQAQLEEFQADIYVIGNVARRGMDVVE 81
   Sbjct: 21 RQLGHEVTGSDANVYPPMSTLLEKQGIELIQGYDASQLEP-QPDLVIIGNAMTRGNPCVE 79

45  Query: 82 AILNRLPYISGPQWLAENVLHHHWVLGVAGTHGKTTASMLAWVLEYAGLAPGFLIGGV 141
   Sbjct: 80 AVLEKNIPYMSGPQWLHDFVLRDRVWLAVAGTHGKTTAGMATWILEQCGYKPGFVIGGV 139

50  Query: 142 PENFGVSARLPQTPRQDPNSKSPFFVIEADEYDTAFFDKRSKFVHYRPRTAVLNNLEFDH 201
   Sbjct: 140 PGNFEVSAHL-----GESDFFVIEADEYDCAFFDKRSKFVHYCPRTLILNNLEFDH 190

55  Query: 202 ADIFADLGAIQTQFHHLVVRTVPSEGLIVCNGQQQSLQDTLDKGCWTPVEKFGTGHGWQIG 261
   Sbjct: 191 ADIFDDLKAIQKQFHHLVRIVPGQGRIIWPENDINLKQTMAMGCWSEQLVGEQGHWQAK 250

60  Query: 262 EVNADGS-FDVLLDGKKAGHVAWDLMGHNRNMALAVIAAARHAGVDVQTACEALGAFKN 320
   Sbjct: 251 KLTTDASEWEVLLDGEKVGKWSLVGEHNMHNLMAIAAARHVGVAADAANALGSGFIN 310

65  Query: 321 VKRRMEIKGTANGITVYDDFAHHPTAIETTIQGLRQRVGG-ARILAVLEPRSNMTKLGTM 379
   Sbjct: 311 ARRLERLGEANGVTVYDDFAHHPTAILATLAALRGKVGCTARIIVLEPRSNMTKMGIC 370

   Query: 380 KSALPASLKEADQVF-CYAGGADWDVAEALAPLGCRLRVGKDFDTFVAEIVKNARTGDHI 438
   Sbjct: 371 KDDLAPSLGRADEVFLLOPAHIPWQVAEVAEACVQPAHWSGDVDTLADMVVKTAQPGDHI 430

   Query: 439 LVMSNGGFGGIHTKLLDAL 457

```

LVMSNGGFGGIH KLLD L
Sbjct: 431 LVMSNGGFGGIHQKLLDGL 449

Based on this analysis, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

- 5 ORF132-1 (26.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 20A shows the results of affinity purification of the His-fusion protein, and Figure 20B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for FACS analysis (Figure 20C) and ELISA (positive result). These
10 experiments confirm that ORF132 is a surface-exposed protein, and that it is a useful immunogen.

Example 103

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 875>

```
1  . . CCGGGCTATT ACGGCTCGGA TGACGAATTT AAGCGGGCAT TCGGAGAAAA
51  CTCGCCGACA TmCAAGAAAC ATTGCAACCG GAGCTGCGGG ATTTATGAAC
15  101  CCGTATTGAA AAAATACGGC AAAAAGCGCG CCAACAACCA TTCGGTCAGC
151  ATTAGTGCGG ACTTCGCGCA TTATTTTCATG CCGTTCGCCA GCTATTCGCG
201  CACACACCGT ATGCCCCAACA TCCAAGAAAT GTATTTTCC CAAATCGGCG
251  ACTCCGGCGT TCACACCGCC TTAACCAGC AGCGCGCAA CACTTGCCAA
301  TTTGGCTTCr ATACCTATAA AAAAGGATTG TTAACAACAG ATGATACATT
20  351  AGGATTAAAA CTGGTCGGCT ACCGCAGCCG CATCGACAAC TACATCCACA
401  ACGTTTACGG GAAATGGTGG GATTGAACG GGGATATTCC GAGCTGGGTC
451  AGCAGCACCG GGCTTGCCTA CACCATCCAA CATCGCrATT TCawAGACAA
501  AGTGCATCAA nnnnnnnnnn nnnnnnnnnn nnnnTACGAT TATGGGCGGT
551  TTTTCACCAA CCTTCTTAC GCCTATCAA AAAGCACGCA ACCGACCAAC
25  601  TTCAGCGATG CGAGCGAATC GCCCAACAAT GCGTCCAAAG AAGACCAACT
651  CAAACAAGGT TATGGGTTGA GCAGGGTTTC CGCCTGCCG CGAGATTACG
701  GACGTTTGGA AGTCGGTACG CGCTGGTTGG GCAACAACT GACTTTGGGC
751  GGCGCGATGC GCTATTTCGG CAAGAGCATC CGCGCGACGG CTGAAGAACG
801  CTATATCGAC GGACCAACG GGGGAAATAC CAGCAATTC CGGCAACTGG
30  851  GCAAGCGTTC CATCAACAA ACCGAAACTC TTGCCCGCCA GCCTTTGATT
901  TTWATTTTtA ACGCGCTTA CGAGCCGAG AAAAACCTTA TTTTCCGCGC
951  CGAAGTCAA AATCTGTTCG ACAGCGGTTA TATCGATCCG CTCGATGCGG
1001 GCAATGATGC GGCAAC .GAG CGTTATTACA GCTCGTTCGA CCCGAAAGAC
1051 AAGGACrrAG ACGTAACGTG TAATGCTGAT AAAACGTTGT GCaACGGCAA
35  1101 ATACGGCGGC ACAAGCAAAA GCGTATTGAC CAATTTTGCA CGCGGACGCA
1151 CCTTTTtGAT GACGATGAGC TACAAGTTTT AA
```

This corresponds to the amino acid sequence <SEQ ID 876; ORF133>:

```
1  . . PGYYGSDDEF KRAFGENSPT XKKHCNRS CG IYEPVLK KYG KKRANNHSVS
51  ISADFGDYFM PFASYSRTHR MPNIQEMYFS QIGDSGVHTA LKPERANTWQ
40  101  FGFTYKKG LKQDDTLGLK LVGYRSRIDN YIHNVYGKWW DLNGDIPSWV
151  SSTGLAYTIQ HRXFXDKVHQ XXXXXXXXYD YGRFFTNLSY AYQKSTQPTN
201  FSDASESPNN ASKEDQLKQG YGLSRVSALP RYGRLEVGT RWLGKLTILG
251  GAMRYFGKSI RATAERYID GTNGGNTSNF RQLGKRSIKQ TETLARQPLI
301  XDFNAAYEPK KNLI FRAEVK NLFDRRYIDP LDAGNDAAXE RYSSFDPKD
45  351  KDXDVT CNAD KTL CNKYGG TSKSVLTNFA RGRTFLMTMS YKF*
```

Further work revealed the further partial DNA sequence <SEQ ID 877>:

```
1  GAGGCGCAGA TACAGGTTTT GGAAGATGTG CACGTCAAGG CGAAGCGCGT
51  ACCGAAAGAC AAAAAGTGT TTACCGATGC GCGTGCCGTA TCGACCCGTC
50  101  AGGATATATT CAAATCCAGC GAAAACCTCG ACAACATCGT ACGCAGCATC
151  CCCGGTGGT TTACACAGCA AGATAAAAGC TCGGGCATTG TGTCTTTGAA
201  TATTCGCGGC GACAGCGGGT TCGGGCGGGT CAATACGATG GTGGACGGCA
251  TCACGCAGAC CTTTATTTCG ACTTCTACCG ATGCGGGCAG GGCAGGCGGT
```


-474-

301 TCATCTCAAT TCGGTGCATC TGTGACAGC AATTTTATTG CCGGACTGGA
 351 TGTCGTCAAA GGCAGCTTCA GCGGCTCGGC AGGCATCAAC AGCCTTGCCG
 401 GTTCGGCGAA TCTGCGGACT TTAGGCGTGG ATGACGTCGT TCAGGGCAAT
 451 AATACCTACG GCCTGTGCT AAAAGGTCTG ACCGGCACC AATCAACCAA
 501 AGGTAATGCG ATGGCGGCGA TAGGTGCGCG CAAATGGCTG GAAAGCGGAG
 551 CATCTGTGCG TGTGCTTTAC GGCACAGCA GCGCGAGCGT GCGCGAAAT
 601 TACCGCGTGG GCGGCGGCGG GCAGCACATC GGAATTTTG GCGCGGAATA
 651 TTTGGAACGG CGCAAGCAGC GATATTTTGT ACAAGAGGGT GCTTTGAAAT
 701 TCAATTCCGA CAGCGGAAAA TGGGAGCGGG ATTTACAAAG GCAACAGTGG
 751 AAATACAAGC CGTATAAAAA TTACAACAAC CAAGAACTAC AAAAAATACAT
 801 CGAAGAGCAT GACAAAAGCT GCGGGGAAAA CCTg.CaCCG CAATACGACA
 851 TTACCCCAT CGATCCGTCC AGCCTGAAGC AGCAGTCGGC AGGCAATCTG
 901 TTTAAATTGG AATACGACGG CGTATTCAAT AAATACACGG CGCAATTTCCG
 951 CGATTTAAAC ACCAAAATCG GCAGCCGCAA AATCATCAAC CGCAATTATC
 1001 AGTTCAATTA CGGTTTGTCT TTGAACCCGT ATACCAACCT CAATCTGACC
 1051 GCAGCCTACA ATTCGGGCAG GCAGAAATAT CCGAAAGGGT CGAAGTTTAC
 1101 AGGCTGGGGG CTTTAAAGG ATTTTGAAAC CTACAACAAC GCGAAAAATCC
 1151 TCGACCTCAA CAACACCGCC ACCTTCCGGC TGCCCGCGA AACCGAGTTG
 1201 CGAACCACCT TGGGCTTCAA TTATTCCAC AACGAATACG GCAAAAACCG
 1251 CTTTCTGAA GAATTGGGGC TGTTTTCGA CGGTCTGAT CAGGACAACG
 1301 GGCTTTATTC CTATTGGGG CGGTTAAGG GCGATAAAG GCTGCTGCCC
 1351 CAAAATCAA CCATTGTCCA ACCGGCCGGC AGCCAATATT TCAACAGCTT
 1401 CTACTTCGAT GCGCGCTCA AAAAGACAT TTACCGCTTA AACTACAGCA
 1451 CCAATACCGT CGGCTACCGT TTCGGCGCG AATATACGGG CTATTACGGC
 1501 TCGGATGACG AATTTAAGCG GGCATTGCGA GAAACTCGC CGACATACAA
 1551 GAAACATTGC AACCGAGCT GCGGGATTTA TGAACCCGTA TTGAAAAAAT
 1601 ACGGCAAAAA GCGCGCCAAC AACCATTGCG TCAGCATTAG TGCGGACTTC
 1651 GCGGATTATT TCATGCCGTT CGCCAGCTAT TCGCGCACAC ACCGTATGCC
 1701 CAACATCCAA GAAATGTATT TTTCCAAAT CGGCGACTCC GCGGTTTACA
 1751 CCGCCTTAAA ACCAGGCGC GCAACACTT GGCAATTGG CTTCAATACC
 1801 TATAAAAAAG GATTGTAAA ACAAGATGAT ACATTAGGAT TAAACTGGT
 1851 CCGCTACCGC AGCCGCATCG ACAACTACAT CCACAACGTT TACGGGAAAT
 1901 GGTGGGATT GAACGGGGAT ATTCGAGCT GGGTGAGCAG CACCGGGCTT
 1951 GCCTACACCA TCCAACATCG CAATTTCAAA GACAAAGTGC ACAAACACGG
 2001 TTTTGAGTTG GAGCTGAATT ACGATTATGG GCGTTTTTTC ACCAACCTTT
 2051 CTTACGCCTA TCAAAAAAGC ACGCAACCGA CCAACTTCAG CGATGCGAGC
 2101 GAATCGCCCA ACAATGCGTC CAAAGAAGAC CAACTCAAAC AAGGTTATGG
 2151 GTTGAGCAGG GTTCCGCCC TGCCGCGAGA TTACGGACGT TTGGAAGTCG
 2201 GTACGCGCTG GTTGGGCAAC AACTGACTT TGGGCGGCGC GATGCGCTAT
 2251 TTCGGCAAGA GCATCCGCGC GACGGCTGAA GAACGCTATA TCGACGGCAC
 2301 CAACGGGGGA AATACAGCA ATTCCGCGA ACTGGGCAAG CGTTCCATCA
 2351 AACAAACCGA AACTCTTGCC CGCCAGCCTT TGATTTTGA TTTTACGCC
 2401 GCTTACGAGC CGAAGAAAA CTTATTTTC CGCGCCGAAG TCAAAAATCT
 2451 GTTCGACAGG CGTTATATCG ATCCGCTCGA TGCGGGCAAT GATGCGGCAA
 2501 CGCAGCGTTA TTACAGCTCG TTCGACCCGA AAGACAAGGA CGAAGACGTA
 2551 ACGTGAATG CTGATAAAAC GTTGTGCAAC GGCAAATACG GCGGCACAAG
 2601 CAAAAGCGTA TTGACCAAT TTGCACGCGG ACGCACCTTT TTGATGACGA
 2651 TGAGCTACAA GTTTTAA

This corresponds to the amino acid sequence <SEQ ID 878; ORF133-1>:

50 1 EAQIQVLEDV HVKAKRVPKD KKVFTDARAV STRQDIFKSS ENLDNIVRSI
 51 PGAFTQDDKS SGIVSLNIRG DSGFGRVNTM VDGITQTFFYS TSTDAGRAGG
 101 SSQFGASVDS NFIAGLDVVK GSFSGSAGIN SLAGSANLRT LGVDDVVOGN
 151 NTYGLLLKGL TGTNSTKGNA MAAIGARKWL ESGASVGVLY GHSRRSVAQN
 201 YRVGGGGQHI GNFGAEYLER RKQRYFVOEG ALKFNSDSGK WERDLQRQQW
 251 KYKPYKNYNN QELQKYIEEH DKSARENLXP QYDITPIDPS SLKQQSAGNL
 301 FKLEYDGVFN KYTAQFRDLN TKIGSRKIIN RNYQFNYGLS LNPYTNLNL
 351 AAYNSGRQKY PKGSKFTGWG LLKDFETYN AKILDNLNTA TFRPLPRET
 401 QTTLGFNYFH NEYGNRFP EELGLFDDGPD QDNGLYSYLG RFGDKGLLP
 451 QKSTIVQPAG SQYFNTFYFD AALKKDIYRL NYSTNTVGYR FGGEYTGYYG
 501 SDDEFKRAFG ENSPTYKKHC NRSCGIYEPV LKKGKGRAN NHSVSIADF
 551 GDYFMPFASY SRTHRMPIQ EMYFSQIGDS GVHTALKPER ANTWFQFNT
 601 YKKGLLKQDD TLGLKLVGYR SRIDNYIHNV YGKWWDLNGD IPSWVSSTGL
 651 AYTIIQHRNF DKVHKHGFEL ELNYDYGRFF TNLAYYQKS TOPTNFSAS
 701 ESPNNASKED QLKQGYGLSR VSALPRDYGR LEVGRWLGN KLTGGAMRY
 751 FGKSIRATAE ERYIDGTNGG NTSNFRQLGK RSIKQETELA RQPLIFDFYA
 801 AYEPRKNLIF RAEVKNLFD RYIDPLDAGN DAATQRYSS FDPKDKDEDV
 851 TCNADKTLN GKYGGTSKSV LTNFARGRTF LMTMSYKF*

Computer analysis of this amino acid sequence gave the following results:

Homology with with the probable TonB-dependent receptor HI121 of *H.influenzae* (accession number U32801)

ORF133 and HI121 show 57% aa identity in 363aa overlap:

```

5  Orf133: 31 IYEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTA 90
    I EP+L K G K+A NHS ++SA+ DYFMPF +YSRTHRMPNIQEM+FSQ+ ++GV+TA
    HI121: 563 INEPILHKS GHKKA FNHSATLSAELSDYFMPFFTYSRTHRMPNIQEMFFSQVSNAGVNTA 622

10 Orf133: 91 LKPERANTWQFGFXYTKKGLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWV 150
    LKPE+++T+Q GF TYKKGL QDD LG+KLVGYRS I NYIHNVYG WW +P+W
    HI121: 623 LKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNYIHNVYGWW--RDGMPTWA 680

15 Orf133: 151 SSTGLAYTIQHRFXDKVHXXXXXXXXXXDYGRFFTNLSYAYQKSTOPTNFSDASESPNN 210
    S G YTI H+ + V YD GRFF N+SYAYQ++ OPTN++DAS PNN
    HI121: 681 ESNGFKYTIHQNYKPIVKKSGVELEINYDMGRFFANVSYAYQRTNQPTNYADASPRPNN 740

20 Orf133: 211 ASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGMARYFGKSIRATAEERYID 270
    AS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTLG A RY+GKS RAT EE YI+
    HI121: 741 ASQEDILKQGYGLSRVSMLPKDYGRLELGTWFDQKLTGLAARYYFGKSKRATIEEYIN 800

25 Orf133: 271 GTNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDP 330
    G+ + R+ ++K+TE + +QP+I D + +YEP K+LI +AEV+NL D+RY+DP
    HI121: 801 GSR-FKKNLTLRENYAYVKKTEDIKKQPIILDHVSYPEIKDLIIKAEVQNLLDKRYVDP 859

    Orf133: 331 LDAGNDAAXERYSSFDPKDKDXDVTNADKTLNCGKYGGTSSKSVLTNFARGRTFLMTMS 390
    LDAGNDAA +RYSS + + C D + C GG+ K+VL NFARGRT+++++
    HI121: 860 LDAGNDAASQRYSSL-----NNSIECAQDSSAC-----GGSDKTVLYNFARGRTYILSLN 910

    Orf133: 391 YKF 393
    YKF
    HI121: 911 YKF 913
30

```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF133 shows 90.8% identity over a 392aa overlap with an ORF (ORF133a) from strain A of *N. meningitidis*:

```

35  orf133.pep          10      20      30
    PGYYGSDDEFKRAFGENSPTXKKHCNRS CGI
    orf133a          FYFDAALKKDIYRLNYSNTNTVG YRFGGX YTGYYXSDDEFKRAFGENSPTXKKHCNRS CGI
    450      460      470      480      490      500

40  orf133.pep          40      50      60      70      80      90
    YEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL
    orf133a          YEPVLKKYGKKRANNHVSISADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL
    510      520      530      540      550      560

45  orf133.pep          100      110      120      130      140      150
    KPERANTWQFGFXYTKKGLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVS
    orf133a          KPERANTWQFGFNTYKKGLKQDDILGLKLVGYRSRIDXYIHNVYGKWWDLNGNIPSWVS
    570      580      590      600      610      620

50  orf133.pep          160      170      180      190      200      210
    STGLAYTIQHRFXDKVHQXXXXXXXXXXDYGRFFTNLSYAYQKSTOPTNFSDASESPNNA
    orf133a          STGLAYTIQHRNFKDKVHKHGFELNLDYXRFFTNLSYAYQKSTOPTNFSDASESPNNA
    630      640      650      660      670      680

55  orf133.pep          220      230      240      250      260      270
    SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGMARYFGKSIRATAEERYIDG
    orf133a          SKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTGGMARYFGKSIRATAEERYIDX
    690      700      710      720      730      740

60  orf133.pep          280      290      300      310      320      330
    TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDPL
    orf133a          TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLIFRAEVKNLFDRRYIDPL
65

```

A partial ORF133a nucleotide sequence <SEQ ID 879> is:

20
25
30
35
40
45
50
55
60
65

1	AAGACAAA	AAGTGTTC	CGATCGCG	GCCGTATCGA	CCCGTCAGGA
51	TATATTCAAA	TCCANCGAAA	ACCTCGACAA	CATCGTACGC	ANCATCCCCG
101	GTGCGGTTAC	ACANCAANAT	AAAAGCGCT	CCGNTTGTGC	TTTGAATATT
151	CGCGCGCAGA	CGGGGTTCCG	CGGGGTCAAT	ACNATGGTNG	ACGGCATCAC
201	NCANACCTTT	TATTCGACTT	CTACCGATGC	GGGCAGGGCA	GGCGGTTTAT
251	CTCAATTCCG	TGCACTCTGT	GACAGCAATT	TTATNGCCGG	ACTGGATGTC
301	GTCAAAGGCA	GCTTCAGCGG	CTCGGCAAGC	ATCAACAGCC	TTGCGGGTTC
351	GGCGAATCTG	CGGACTTTAN	CGGTGGATGA	TGTCTTTCAG	GGCAATANTA
401	CNTACGGCTG	GCTGCTAAAA	GGTCTGACCG	GCACCAATTG	AACCAAAGGT
451	AATGCGATGT	CGCGGATAGG	TGCGCGCAAA	TGGCTGGAAA	CGCGAGCATC
501	TGTCCGGTGT	CTTTACGGCG	ACAGCAGGCG	CAGCGTGGCG	CAAAATTACC
551	CGGTGGGCGG	CGGCGGGCAG	CACATCGGAA	ATTTTGGCGC	GGAATATCTG
601	GAACGACGCA	AGCAACGATA	TTTTGAGCAA	CGAAGGCGGT	TGAAATTCAA
651	TTCCAAACAG	TGAAAAATGG	AGCGGGATT	CCAAAGATCG	TACTGGAAAA
701	CCAAGTGGTA	TCAAAAATAC	GATGCCCCCC	AAGAACTGCA	AAAATACATC
751	GAAGGTCATG	ATAAAAGCTG	CGGGGAAAC	CTGGCGCCGC	AATACGACAT
801	CACCCCATC	GATCCGCTCA	GCCTGAAGCN	CGAGTCGGCA	GGCAACCTGT
851	TTAAATTGGA	ATACGACGCG	GTATTCAATA	AATACACGCG	GCAATTTCCG
901	GATTTAAACA	CCAAATCCGG	CAGCCGCAAA	ATCATCAACC	GCAATTATCA
951	ATTCATATT	GGTTTGTCTT	TGAACCCGTA	TACCAACCTC	AATCTGACCG
1001	CAGCCTACAA	TTCCGGCAGC	CAGAAATATC	CGAAGGGTCT	GAAGTTTACA
1051	GGCTGGGGGC	TTTTNAAAAG	TTTTGAAACC	TACAAACAAC	CAAAAATCCT
1101	CGACCTCANC	AACACCTCCA	CCTTCCGGCT	GCCCCGTGAA	ACCGAGTTGC
1151	AAACCACTTT	GGGCTTCAAT	TATTTCCACA	ACGAATACCG	CAAAAACCCG
1201	TTTCTGAAG	AATTGGGGCT	GGTTTTCGAC	GGTCCGGATC	ANGACAACGG
1251	GCTTTATTTC	TATTTGGGGC	GGTTTAAGGG	CGATAAAGGG	CTGCTGCCCC
1301	AAAAATCAAC	CATTGTCCAA	CCGCGCCGGA	GCCATATATT	CAACACGTTT
1351	TACTTCGATG	CCGCGCTCAA	AAAAGACATT	TACCGCTTAA	ACTACAGCAC
1401	CAATACGCTC	GGCTACCGTT	TCCGCGGCNA	ATATACGGGC	TATTACNGCT
1451	CGGATGACGA	ATTTAAGCGG	GCATTCCGAG	AAAACCTCGC	GACATACANG
1501	AAACATTGCA	ACGAGAGCTG	CGGAATTTAT	GAACCCGTAT	TGAAAAATAA
1551	CGGCAAAAAG	CGCGCAACGA	ACCATTCGGT	CAGCATATAGT	CGCGACTTCG
1601	GCGATTATTT	CATGCCGTTT	GCCAGCTATT	CGCGCACACA	CCGTATGCCC
1651	AACATCCAAG	AAATGTATTT	TTCCCAAATC	GGCGACTCCG	TTCCTCACAC
1701	CGCCTTAAAA	CCAGAGCGCG	CAACACCTTG	GCAATTTGGC	TTCAATACCT
1751	ATAAAAAAGG	ATTGTTTAAA	CAAGATGATA	TATTAGGATT	AAAACCTGGT
1801	GGTACCAGCA	GCCGATTCGA	CNACTACATC	CACAACGTTT	ACGGGAAATG
1851	GTGGGATTTG	AACGGGAATA	TTCCGAGCTG	GGTCAGCAGC	ACCGGGCTTG
1901	CCTACACCAT	CAACACCCGC	AATTTCAAAG	ACAAAGTGCA	CAACACCGGT
1951	TTTGAGTTGG	AGCTGAATTA	CGATTATNGG	CGTTTTTTCA	CCAACCTTTC
2001	TTACGCCTAT	CAAAAAAGCA	CGCAACCCAG	CAACTTCAGC	GATCGGAGCG
2051	AATCGCCCAA	CAATGCGTCC	AAAGAAGACC	AACTCAACA	AGGTTATTGG
2101	TTGAGCAGGG	TTTCCGCCCT	GCCGCGAGAT	TACGGACGTT	TGGAAGTCGG
2151	TACGCGCTGG	TTGGGCAACA	AACTGACTTT	GGGCGGCGCG	ATGCGCTATT
2201	TCGGCAAGAG	CATCCGCGCG	ACGGCTGAAG	AACGCTATAT	CGACGNCACC
2251	AATGGGGNAN	NTACAGCAAA	TTTCCGGCAA	CTGGGCAAGC	GTTCATCAN
2301	ACAAACCGAA	ACCCTTGCCC	GCCAGCCTTT	GATTTTTGAT	TNTACGCCG
2351	CTTACGAGCC	GAAGAAAAAN	CTTATTTTCC	GCGCCGAAGT	CAAAATCTCT
2401	TTGACAGAGC	GTATTATCGA	TCCGCTCGAT	CGCGGCAATG	ATCGGCAAC
2451	GCAGCGTTAT	TACAGTTCGT	TCGACCCGAA	AGACAAGGAC	GAAGAAGTAA
2501	CGTGTAAATG	TGATAACACG	TTATGCAACG	GCAAATACGG	CGGCACAAGC
2551	AAAAGCGTAT	TGACCAATTT	TGCACGCGGA	CNCACCTTTT	TGATAACGAT
2601	GAGCTCAAG	TTTTTAA			

This encodes a protein having (partial) amino acid sequence <SEQ ID 880>:

```

      1 KDKKVFTDAR AVSTRQDIFK SXENLDNIVR XIPGFTXQX KSSGXVSLNI
      51 RXDSGFGRVN TMVDGITXTF YSTSTDAGRA GGSSQFGASV DSNFXAGLDV
    101 VKGSFSGSAG INSLAGSANL RTLXVDDVVQ GNXTYGLLLK GLTGTNSTKG
    151 NAMAAGARK WLESGASVGV LYGHSRRSVA QNYRVGGGGQ HIGNFGAEYL
    201 ERRKQRYFEQ EGGLKFNSNS GKWERDFQKS YWKTWYQKY DAPQELQKYI
    251 EGHDKSWREN LAPQYDITPI DPSSLKXQSA GNLFKLEYDG VFNKYTAQFR
    301 DLNTKIGSRK IINRNYQFNY GLSLNPTYNL NLTAAYNSGR QKYPKGSKFT
    351 GWGLXKDFET YNNAKILDIX NTSTFRLPRE TELQTTLGFN YFHNEYGKNR
    401 FPEELGLFFD GPDNDGLYS YLGRFKGDKG LLPQKSTIVQ PAGSQYFNTF
    451 YFDAALKKDI YRLNYSTNTV GYRFGGXVTG YYXSDDEFKR AFGENSPTYX
    501 KHCNQSCGIY EPVLKKYGKK RANNHSVSIS ADFGDYFMPF ASYSRTHRMP
    551 NIQEMYFSQI GDSGVHTALK PERANTWQFG FNTYKKGILL QDDILGLKLV
    601 GYRSRIDXYI HNVYGKWDDL NGNIPSWVSS TGLAYTIQHR NFKDKVHKHG
    651 FELELNIDYX RFFTNLSYAY QKSTQPTNFS DASESPNNAS KEDQLKQYGG
    701 LSRVSALPRD YGRLEVGTWR LGNKLTLGGA MRYFGKSIRA TAEERYIDXT
    751 NGXXTSNFRQ LGKRSIXQTE TLARQPLIFD XYAAYPEKKX LIFRAEVKNL
    801 FDRRYIDPLD AGNDAATQRY YSSFDPKDKD EEVTCNDNDT LCNGKYGGTS
    851 KSVLTNFARG XTFLITMSYK F*
  
```

20 ORF133a and ORF133-1 show 94.3% identity in 871 aa overlap:

```

                                10      20      30      40
orfl33a.pep                    KDKKVFTDARAVSTRQDIFKSXENLDNIVRXIPGFTXQXKS
                                |||||
orfl33-1                        EAQIQVLEDVHVAKRVPKDKKVFTDARAVSTRQDIFKSSENLDNIVRSIPGFTQODKS
                                10      20      30      40      50      60

                                50      60      70      80      90      100
orfl33a.pep                    SGXVSLNIRXDSGFGRVNTMVDGITXTFYSTSTDAGRAGSSQFGASVDSNFXAGLDVVK
|| |||||
orfl33-1                        SGIVSLNIRGDSGFGRVNTMVDGITQTFYSTSTDAGRAGSSQFGASVDSNFIAGLDVVK
                                70      80      90      100      110      120

                                110     120     130     140     150     160
orfl33a.pep                    GSFSGSAGINSLAGSANLRTLXVDDVVQGNXTYGLLLKGLTGTNSTKGNAMAAGARKWL
|| |||||
orfl33-1                        GSFSGSAGINSLAGSANLRTLGVDDVVQGNNTYGLLLKGLTGTNSTKGNAMAAGARKWL
                                130     140     150     160     170     180

                                170     180     190     200     210     220
orfl33a.pep                    ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFEQEGGLKFNSNSGK
|| |||||
orfl33-1                        ESGASVGVLYGHSRRSVAQNYRVGGGGQHIGNFGAEYLERRKQRYFVQEGALKFNSDSGK
                                190     200     210     220     230     240

                                230     240     250     260     270     280
orfl33a.pep                    WERDFQKSYWKTWYQKYDAPQELQKYIEGHDKSWRENLAPOYDITPIDSSLKXQSAGN
|| |||||
orfl33-1                        WERDLQRQWQKYKPYKNYN-QELQKYIEHDKSWRENLPQYDITPIDSSLKQQSAGN
                                250     260     270     280     290

                                290     300     310     320     330     340
orfl33a.pep                    LFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNYGLSLNPTYNLNLTAAYNSGRQK
|| |||||
orfl33-1                        LFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNYGLSLNPTYNLNLTAAYNSGRQK
                                300     310     320     330     340     350

                                350     360     370     380     390     400
orfl33a.pep                    YPKGSKFTGWGLXKDFETYNNAKILDIXNTSTFRLPRETELQTTLGFNHYFHNEYGKNRFP
|| |||||
orfl33-1                        YPKGSKFTGWGLXKDFETYNNAKILDIXNTSTFRLPRETELQTTLGFNHYFHNEYGKNRFP
                                360     370     380     390     400     410

                                410     420     430     440     450     460
orfl33a.pep                    EELGLFFDGPDXNDGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
|| |||||
orfl33-1                        EELGLFFDGPQDNDGLYSYLGRFKGDKGLLPQKSTIVQPAGSQYFNTFYFDAALKKDIYR
                                420     430     440     450     460     470
  
```

```

                                470      480      490      500      510      520
orf133a.pep  LNYSTNTVGYRFGGXYTGYYXSDDEFKRAFGENSPTYKKHCNQSCGIYEPVLKKYGKKRA
              |||||
orf133-1     LNYSTNTVGYRFGGEYTGYYGSDDEFKRAFGENSPTYKKHCNRSCGIYEPVLKKYGKKRA
              480      490      500      510      520      530

```

10

orf133a.pep
 NNHSVSISADFGDYFMPFASYSRTHRPNIQEMYFSQIGDSGVHTALKPERANTWQFGFN
 |||||
 orf133-1
 NNHSVSISADFGDYFMPFASYSRTHRPNIQEMYFSQIGDSGVHTALKPERANTWQFGFN
 540 550 560 570 580 590

15

```

      590      600      610      620      630      640
orf133a.pep  TYKKGLLKQDDILGLKLVGYRSRIDXYIHNVYKQWDLNGNIPSWVSSTGLAYTIQHRNF
              |||||
orf133-1     TYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYKQWDLNGDIPSWVSSTGLAYTIQHRNF
      600      610      620      630      640      650

```

20

```

      650      660      670      680      690      700
orf133a.pep  KDKVHKHGFLELNLDYXRFFTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS
              |||||
orf133-1     KDKVHKHGFLELNLDYGRFFTNLSYAYQKSTQPTNFSDAESPNNASKEDQLKQGYGLS
      660      670      680      690      700      710

```

25

```

              710      720      730      740      750      760
orf133a.pep  RVSA LPRDYGRLEV GTRWLG NKLTLGGAMRYFGKSI RATAEERYIDXTNGX XTSNFRQLG
              |||||
orf133-1     RVSA LPRDYGRLEV GTRWLG NKLTLGGAMRYFGKSI RATAEERYIDGTNGGNTSNFRQLG
              720      730      740      750      760      770

```

30

```

              770      780      790      800      810      820
orf133a.pep  KRSIXQTETLARQPLIFDXYAAEYPPKKXLI FRAEVKNLFDRRYIDPLDAGNDAATQRYYS
              ||||| ||||| ||||| ||||| ||||| |||||
orf133-1     KRSIXQTETLARQPLIFDFYAAEYPPKKNLI FRAEVKNLFDRRYIDPLDAGNDAATQRYYS
              780      790      800      810      820      830

```

35

```

      830      840      850      860      870
orf133a.pep  SFDPKDKDEEVTCNDDNTLCNGKYGGTSKSVLTNFARGXTFLITMSYKFX
              |||||:|||||:|||||:|||||:|||||:|||||:|||||
orf133-1     SFDPKDKDEDVTCNADKTLCNGKYGGTSKSVLTNFARGRTFLMTMSYKFX
      840      850      860      870      880

```

ORF133 shows 92.3% identity over 392 aa overlap with a predicted ORF (ORF133ng) from *N.*

45 gonorrhoeae:

```
orf133.pep          PGYYGSDDEFKRAFGENSPTXKKHCNRSCGI      31
```

orf133ng FYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKRAFGENSPAYKEHCDPSCGL 560

50

orf133.pep YEPVLKKYGGKRRANNHSVSIADFGDYFMPFASYSRTHRMPNIQEMYFSQIGDSGVHTAL 91

orf133nq YEPVLKKYGGKRRANNHVSVISADFGDYFMPFAGYSRTHRMPNIQEMYFSQIGDSGVHTAL 620

55

orf133.pep KPERANTWQFGFXTYKKGLLKQDDTLGLKLVGYRSRIDNYIHNVYGKWWDLNGDIPSWVS 151

orf133ng KPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYIHNVYKGWWDLNGDIPSWVG 680

orf133.pep STGLAYTIQHRXFXDKVHQXXXXXXXXXYDYGRFTNLSYAYQKSTQPTNFSDASESPNNA 211

60

orf133ng STGLAYTIRHRNFKDKVHKHGFLELNYDYGRFTNLISYAYQKSTQPTNFSDASESPNNA 740

orf133.pep SKEDQLKQGYGLSRVSALPRDYGRLEVGT RWLG NKLT LGGAMRYFGKSIRATAEERYIDG 271

orf133ng SKEDQLKQGYLSRVSALPRDYGRLVGTRWLGNKLTGGAMRYFGKSIRATAEERYIDG 800

65

orf133.pep TNGGNTSNFRQLGKRSIKQTETLARQPLIXDFNAAYEPKKNLI FRAEVKNLFDRRYIDPL 331

orf133ng TGGGNTSNVRQLGKRSIKOTETLAROPLI FDFYAAYPEKKNLIFRAEVKNLFD RRYIDPL 860

orf133.pep DAGNDAAXERYSSFDPKDKDXDVTCTNADKTLNCGKYGGTSKSVLTNFARGRTFLMTMSY 391
 |||||:::||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 orf133ng DAGNDAATQRYSSFDPKDKDEDVTCTNADKTLNCGKYGGTSKSVLTNFARGRTFLMTMSY 920
 5 orf133.pep KF 393
 ||
 orf133ng KF 922

The complete length ORF133ng nucleotide sequence <SEQ ID 881> is predicted to encode a
 10 protein having amino acid sequence <SEQ ID 882>:

1 MRSSFRLKPI CFYLMGVMLY HHSYAEDAGR AGSEAQIQVL EDVHVKAQRV
 51 PKDKKVFDA RAVSTRQDVF KSGENLDNIV RSIPGAFTQQ DKSSGIVSLN
 101 IRGDSGFGRV NTMVDGITQT FYSTSTDAGR AGSSSQFGAS VDSNFIAGLD
 151 VVKGSFSGSA GINSLAGSAN LRTLGVDDVV QGNNTYGLLL KGLTGTNSTK
 201 GNAMAAIGAR KWLESGASVG VLYGHSRRGV AQNYRVGGGG QHIGNFGEEY
 251 LERRKQYFV QEGGLKFNAG SGKWERDLQR QYWKTKWYKK YEDPQELQKY
 301 IEEHDKSWRE NLAPQYDITP IDPSGLKQQS AGNLLNLEYD GFENKYTAQF
 351 RDLNTRIGSR KIINRNYQFN YGLSLNPYTN LNLTAAYNSG RQKYPKGAKF
 401 TGVGLLKDFE TYNNAKILD NNTATFRLPR ETELQTTLGF NYFHNEYGKN
 451 RFPEELGLFF DGPDQDNGLY SYLGRFKGDK GLLPQKSTIV QPAGSQYFNT
 501 FYFDAALKKD IYRLNYSNA INYRFGGEYT GYYGSENEFK RAFGENSPAY
 551 KEHCDPSCGL YEPVLKKYK KRANNHSVSI SADFGDYFMP FAGYSRTHRM
 601 PNIQEMYFSQ IGDGSGVHTAL KPERANTWQF GFNTYKKGLL KQDDILGLKL
 651 VGYRSRIDNY IHNVYKQWD LMGDIPSWVG STGLAYTIRH RNFKDKVHKH
 701 GFELELNDY GRFFTNLSYA YOKSTQPTNF SDASESPNNA SKEDQLKQGY
 751 GLSRVSALPR DYGRLEVGT WLGNKLTGG AMRYFGKSIR ATAERYIDG
 801 TNGGNTSNVR QLGKRSIKQT ETLARQPLIF DFYAAEYPPK NLI FRAEVKN
 851 LFDRLYIDPL DAGNDAATQ YSSFDPKDK DEDVTCTNADK TNCNGKYGGT
 901 SKSVLTNFAR GRTFLMTMSY KF*

30 A variant was also identified, being encoded by the gonococcal DNA sequence <SEQ ID 883>:

1 ATGAGATCTT CTTTCCGTT GAAGCCGATT TGTTTTATC TTATGGGTGT
 51 TATGCTATAT CATCATAGTT ATGCCGAAGA TGCAGGGCGC GCGGGCAGCG
 101 AGGCGCAGAT ACAGGTTTTG GAAGATGTGC ACGTCAAGGC GAAGCGCGTA
 151 CCGAAAGACA AAAAAGTGT TACCGATGCG CGTGCCGTAT CGACCCGTca
 201 gGATGTGTT AAATCCGGCG AAAACCTCGA CAACATCGTA CGCAGCATAC
 251 CCGGTGCGTT TACACAGCAA GATAAAAGCT CGGGCATTGT GTCCTTGAAT
 301 ATTCGCGGCG ACAGCGGGTT CGGGCGGGTC AATACGATGG TGGACGGCAT
 351 CACGCAGACC TTTTATTCGA CTTCTACCGA TGCAGGCGAG GCAGGCGGTT
 401 CATCTCAATT CGGTGCATCT GTCGACAGCA ATTTTATTGC CGGACTGGAT
 451 GTCGTCAAAG GCAGCTTCAG CGGCTCGGCA GGCATCAACA GCCTTGCCGG
 501 TTCGGCGAAT CTGCGGACTT TAGGCGTGGA TGACGTCGTT CAGGGCAATA
 551 ATACCTACGG CCTGCTGCTA AAAGTCTGA CCGGCACCAA TTCAACCAAA
 601 GGTAATGCGA TGGCGGCGAT AGGTGCGCGC AAATGGCTGG AAAGCGGAGC
 651 GTCTGTGCGT GTGCTTTACG GGCACAGCAG GCGCGCGGTG GCGCAAAATT
 701 ACCGCGTGGG CGGCGGCGGG CAGCACATCG GAAATTTTGG TGAAGAAAT
 751 CTGGAACGGC GCAAACAGCA ATATTTTGTA CAAGAGGGTG GTTTGAAATT
 801 CAATGCCGGC AGCGGAAAAT GGAACGGGA TTTGCAAAGG CAATACTGGA
 851 AAACAAAGTG GTATAAAAAA TACGAAGACC CCCAAGAACT GCAAAAATAC
 901 ATCGAAGAGC ATGATAAAAG CTGGCGGGAA AACCTGGCGC CGCAATACGA
 951 CATCACCCCC ATCGATCCGT CCGGCCTGAA GCAGCAGTCG GCAGGCAATC
 1001 TGTTTAAATT GGAATACGAC GCGGTATTCA ATAAATACAC GGCGCAATTT
 1051 CGCGATTAA ACACCAGAA CGGCAGCCG AAAATCATCA ACCGCAATTA
 1101 TCAATTCAAT TACGGTTTGT CTTTGAACCC GTATACCAAC CTAATCTGA
 1151 CCGCAGCCTA CAATTGCGGC AGGCAGAAAT ATCCGAAAGG GGCGAAGTTT
 1201 ACAGGCTGGG GGCTTTTAAA AGATTTTGAA ACCTACAACA ACAGGAAAAT
 1251 CCTCGACCTC AACAACACCG CCACCTCCG GCTGCCCGCG GAAACCGAGT
 1301 TGCAAACCA TTTGGGCTTC AATTATTCC ACAACGAATA CGGCAAAAAC
 1351 CGCTTTCCTG AAGAATTGGG GCTGTTTTTC GACGGTCTGT ATCAGGACAA
 1401 CGGGCTTTAT TCCTATTGCG GCGGTTTAA GGGCGATAAA GGCTGTTGC
 1451 CTCAAAAATC AACCATTGTC CAACCGGCGG GCAGCCAATA TTTCAACACG
 1501 TTCTACTTCG ATGCCGCGCT CAAAAAGAC ATTTACCGCT TAAACTACAG
 1551 CACCAATGCA ATCAACTACC GTTTCGGCGG CGAATATACG GGCTATTACG
 1601 GCTCGGAAA CGAATTTAAG CGGGCATTCT GAGAAAATCT GCCGGCATA
 1651 AAGGAACATT GCGACCCGAG CTGCGGGCTT TATGAACCCG TATTGAAAAA
 1701 ATACGGCAAA AAGCGCGCCA ACAACCATTC GGTGAGCATT AGTGGCGACT
 1751 TCGGCGATTA TTTATGCGC TTCGCGGCT ATTCGCGCAC ACACCGTATG

1801 CCCAACATCC AAGAAATGTA TTTTCCCAA ATCGGCGACT CCGGCGTTCA
 1851 CACCGCCTTA AAACCAGAGC GCGCAAACAC TTGGCAATTT GGCTTCAATA
 1901 CCTATAAAAA AGGATTGTTA AAACAAGATG ATATATTAGG ATTGAAACTG
 1951 GTCGGCTACC GCAGCCGCAT TGACAACACT ATCCACAACG TTTACGGGAA
 5 2001 ATGGTGGGAT TTGAACGGGG ATATTCCGAG CTGGGTCCGGC AGCACC GGCG
 2051 TTGCCTACAC CATCCGACAC CGCAATTTCA AAGACAAAGT GCACAAACAC
 2101 GGTTTTGAGC TGGAGCTGAA TTACGATTAT GGGCGTTTTT TCACCAACCT
 2151 TTCTTACGCC TATCAAAAAA GCACGCAACC GACCAATTTT AGCGATGCGA
 10 2201 GCGAATCGCC CAACAATGCC tccaaAGAAG ACCAACTCAA ACAAGGTTAT
 2251 GGGCTGAGCA GGGTTTCCGC CCTGCCGCGA GATTACGGAC GTTTGGAAGT
 2301 CCGTACGCGC TGGTTGGGCA ACAAACTGAC TTTGGGCGGC GCGAtgcGCT
 2351 ATTTCCGGCA GAGCATCCGC GCGACGGCTG AAGAACGCTA TATCGACGGC
 2401 ACCAACGGGG GAAATACCAG CAATGTCCGG CAACTGGGCA AGCGTTCCAT
 2451 CAAACAAACC GAAACCCCTG CCCGACAGCC TTTGATTTTT GATTTTTACG
 15 2501 CCGCTTACGA GCCGAAGAAA AACCTTATTT TCCGCGCCGA AGTCAAAAC
 2551 CTGTTGACA GGCCTTATAT CGATCCGCTC GATGCGGGCA ATGATGCGGC
 2601 AACGCAGCGT TATTACAGCT CGTTCGACCC GAAAGACAAG GACGAAGACG
 2651 TAACGTGTAA TGCTGATAAA ACGTTGTGCA ACGGCAAATA CCGCGGCACA
 2701 AGCAAAAGCG TATTGACCAA TTTCGCACGC GGACGCACCT TCTTGATGAC
 20 2751 GATGAGCTAC AAGTTTTAA

This corresponds to the amino acid sequence <SEQ ID 884; ORF133ng-1>:

1 MRSSFRLKPI CFYLMGVMLY HHSYAEDAGR AGSEAQIQVL EDVHVAKRV
 51 PKDKKVFTDA RAVSTRQDVF KSGENLDNIV RSIPGAFTQQ DKSSGIVSLN
 101 IRGDSGFGRV NTMVDGITOT FYSTSDAGR AGSSSQFGAS VDSNFIAGLD
 151 VVKGSFSGSA GINSLAGSAN LRTLGVDDVV QGNNTYGLLL KGLTGNSTK
 201 GNAMAAIGAR KWLESGASVG VLYGHSRRGV AQNYRVGGGG QHIGNFGEEY
 251 LERRKQQYFV QEGGLKFNAG SGKWERDLQR QYWKTKWYKK YEDPQELQKY
 301 IEEHDKSWRE NLAPQYDITP IDPSGLKQQS AGNLFKLEYD GVFNKYTAQF
 351 RDLNTRIGSR KIINRNYQFN YGLSLNPYTN LNLTAAYNSG RQKYPKGAKE
 401 TGWGLLKDFE TYNNAKILD NNTATFRLPR ETELQTTLGF NYFHNKYGKN
 451 RFPEELGLFF DGPDPQDNGLY SYLGRFKGDK GLLPKKSTIV QPAGSQYFNT
 501 FYFDAALKKD IYRLNYSYTA INYRFGGEYT GYGSSENEFK RAFGENSPAY
 551 KEHCDPSCGL YEPVLKKGK KRANNHVSIS SADFGDYFMP FAGYSRTHRM
 601 PNIQEMYFSQ IGDSGVHTAL KPERANTWQF GFNTYKKGLL KQDDILGLKL
 651 VGYRSRIDNY IHNVYKQWD LMGDIPSWVG STGLAYTIRH RNFKDKVHHK
 701 FYFELELNYDY GRFFTNLSYA YQKSTOPTNF SDASESPNNA SKEDQLKQGY
 751 GLSRVSALPR DYGRLEVGTG WLGKLTLLGG AMRYFGKSIR ATAEEYIDG
 801 TNGGNTSNVR QLGKRSIKQT ETLARQPLIF DFYAAEPPK NLIFFRAEVKN
 851 LFDRIYIDPL DAGNDAATQR YYSSFDPKDK DEDVTCNADK TLCNGKYGGT
 901 SKSVLTNFAF GRTFLMTMSY KF*

ORF133ng-1 and ORF133-1 show 96.2% identity in 889 aa overlap:

10 20 30 40 50 60
 orf133ng-1.pep SFRLLKPICFYLMGVMLYHHSYAEDAGRAGSEAQIQVLE DVHVAKRVPKDKKVFTDARAV
 45 orf133-1 EAQIQVLE DVHVAKRVPKDKKVFTDARAV
 10 20 30
 70 80 90 100 110 120
 orf133ng-1.pep STRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYS
 50 orf133-1 STRQDIFKSSENLDNIVRSIPGAFTQQDKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYS
 40 50 60 70 80 90
 130 140 150 160 170 180
 orf133ng-1.pep TSTDAGRAGSSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN
 55 orf133-1 TSTDAGRAGSSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVQGN
 100 110 120 130 140 150
 190 200 210 220 230 240
 orf133ng-1.pep NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLYGHSRRGVAQNYRVGGGGQHI
 60 orf133-1 NTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGVLYGHSRRGVAQNYRVGGGGQHI
 160 170 180 190 200 210
 250 260 270 280 290 300
 orf133ng-1.pep GNFGEEYLERRKQYFVQEGGLKFNAGSGKWERDLQRQYWKTKWYKKYEDPQELQKYIEE

-481-

```

||||| |||||||:|||||:|||||: ||||||| || |||: ||| |||||
orf133-1      GNFGAEYLERRKQRYFVQEGALKFNSDSGKWERDLQRQQWKYKPKYNNN-QELQKYIEE
                220      230      240      250      260

5
                310      320      330      340      350      360
orf133ng-1.pep HDKSWRENLAPOYDITPIDPSGLKQKSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII
                |||||||:|||||:|||||: ||||||| || |||: ||| |||||
orf133-1      HDKSWRENLAPOYDITPIDPSGLKQKSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII
                270      280      290      300      310      320

10
                370      380      390      400      410      420
orf133ng-1.pep NRNYQFNYGLSLNPYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDNLNT
                |||||||:|||||:|||||: ||||||| || |||: ||| |||||
orf133-1      NRNYQFNYGLSLNPYTNLNLTAAYNSGRQKYPKGAKFTGWGLLKDFETYNNAKILDNLNT
                330      340      350      360      370      380

15
                430      440      450      460      470      480
orf133ng-1.pep ATFRLPRETELQTTLGFNYPHNEYGKNRFPEELGLFFDGPQDNGLYSYLGRFGDKGGLL
                |||||||:|||||:|||||: ||||||| || |||: ||| |||||
orf133-1      ATFRLPRETELQTTLGFNYPHNEYGKNRFPEELGLFFDGPQDNGLYSYLGRFGDKGGLL
                390      400      410      420      430      440

20
                490      500      510      520      530      540
orf133ng-1.pep POKSTIVQAPAGSQYFNTFYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKRAF
                |||||||:|||||:|||||: ||||||| || |||: ||| |||||
orf133-1      POKSTIVQAPAGSQYFNTFYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSDDEFKRAF
                450      460      470      480      490      500

25
                550      560      570      580      590      600
orf133ng-1.pep GENSPAYKEHCDPSCGLYEPVLKKGKRRANHSVSIADFGDYFMPFAGYSRTHRMPNI
                |||||||:|||||:|||||: ||||||| || |||: ||| |||||
orf133-1      GENSPAYKEHCDPSCGLYEPVLKKGKRRANHSVSIADFGDYFMPFAGYSRTHRMPNI
                510      520      530      540      550      560

30
                610      620      630      640      650      660
orf133ng-1.pep QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYIHN
                |||||||:|||||:|||||: ||||||| || |||: ||| |||||
orf133-1      QEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYIHN
                570      580      590      600      610      620

35
                670      680      690      700      710      720
orf133ng-1.pep VYGKWWDLNGDIPSWVGSTGLAYTIHRNFKDKVHKHGFELNLDYGRFFTNLSYAYQK
                |||||||:|||||:|||||: ||||||| || |||: ||| |||||
orf133-1      VYGKWWDLNGDIPSWVGSTGLAYTIHRNFKDKVHKHGFELNLDYGRFFTNLSYAYQK
                630      640      650      660      670      680

40
                730      740      750      760      770      780
orf133ng-1.pep STQPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWRWLNKLTGGAMR
                |||||||:|||||:|||||: ||||||| || |||: ||| |||||
orf133-1      STQPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWRWLNKLTGGAMR
                690      700      710      720      730      740

45
                790      800      810      820      830      840
orf133ng-1.pep YFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDFYAAAYEPKKNLI
                |||||||:|||||:|||||: ||||||| || |||: ||| |||||
orf133-1      YFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDFYAAAYEPKKNLI
                750      760      770      780      790      800

50
                850      860      870      880      890      900
orf133ng-1.pep FRAEVKNLFDRRYIDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLNKGKYGSTSKS
                |||||||:|||||:|||||: ||||||| || |||: ||| |||||
orf133-1      FRAEVKNLFDRRYIDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLNKGKYGSTSKS
                810      820      830      840      850      860

60
                910      920
orf133ng-1.pep VLTNFARGRTFLMTMSYKFX
                |||||||:|||||:|||||: ||||||| || |||: ||| |||||
orf133-1      VLTNFARGRTFLMTMSYKFX
                870      880

```

70 In addition, ORF133ng-1 is homologous to a TonB-dependent receptor in *H. influenzae*:

sp|P45114|YC17_HAEIN PROBABLE TONB-DEPENDENT RECEPTOR HI1217 PRECURSOR
>gi|1075372|pir||G64110 transferrin binding protein 1 precursor (tbpl) homolog -
Haemophilus influenzae (strain Rd KW20) >gi|1574147 (U32801) transferrin binding
protein 1 precursor (tbpl) [Haemophilus influenzae] Length = 913
Score = 930 bits (2377), Expect = 0.0
Identities = 476/921 (51%), Positives = 619/921 (66%), Gaps = 72/921 (7%)

5
Query: 38 QVLEDVHVAKARVPKDKKVFTDARAVSTRQDVFKSGENLDNIVRSIPGAFTQQDKSSGIV 97
+ L + V K + DKK FT+A+A STR++VFK + +D ++RSIPGAFTQQDK SG+V
10
Sbjct: 29 ETLGQIDVVEKVISNDKKPFTEAKAKSTRENVFKETQTIDQVIRSIPGAFTQQDKSGSVV 88

Query: 98 SLNIRGDSGFGVRNTMVDGITQTFYSTSTDAGRAGGSSQFGASVDSNFIAGLDVVKGSFS 157
S+NIRG++G GRVNTMVDG+TQTFYST+ D+G++GGSSQFGA++D NFIAG+DV K +FS
Sbjct: 89 SVNIRGENGLGRVNTMVDGVTQTFYSTALDSGQSGGSSQFGAIDPNFIAGVDVKNKSNFS 148

15
Query: 158 GSAGINSLAGSANLRTLGVDDVQXXXXXXXXXXXXXXXXXXXXXAMAAIGARKWLESGA 217
G++GIN+LAGSAN RTLG+DV+ M RKWL++G
Sbjct: 149 GASGINALAGSANFRTLGVNDVITDDKPFGIILKGMTGSNATKSNFMTMAAGRWLDNGG 208

20
Query: 218 SVGVLYGHSRRGVAQNYRVGGGGQHIGNFGEEYLERRKQYFVQEGGLKFNAGSGKWERD 277
VGV+YG+S+R V+Q+YR+ GGG+ + + G++ L + K+ YF + G N G+W D
Sbjct: 209 YGVVYGYVSQREVSQDYRI-GGGERLASLGQDILAKEKEAYF-RNAGYILNP-EGQWTPD 265

25
Query: 278 LQRQYWK-----TKWY-----KKYEDPQELQK---YIEE 303
L +++W +Y KK +D ++LQK IEE
Sbjct: 266 LSKKHWSNCNPDYQKNGDCSYRIGSAAKTRREILQELLTNGKKPKDIEKLQKNGDIEE 325

30
Query: 304 HDKSWRENLAPOYDITPIDPSGLKQSSAGNLFKLEYDGVFNKYTAQFRDLNTRIGSRKII 363
DKS+ N QY + PI+P L+ +S +L K EY AQ R L+ +IGSRKI
Sbjct: 326 TDKSFERN-KDQYSVAPIEPGSLQSRSRSHLLKFEYGDHQNLAGAQLRTLNDKIGSRKIE 384

35
Query: 364 NRNYQFNGLSLNYPYTNLNLTAAYNSGRQKYPKGAFTGWGLLKDFETYNNAKILDNLNT 423
NRNYQ NY + N Y +LNL AA+N G+ YPKG F GW + T N A I+D+NN+
Sbjct: 385 NRNYQVNYNFNNNSYLDNLMAAHNIGKTIYPKGGFFAGWQVADKLITKNVANIVDINNS 444

40
Query: 424 ATFRLPRETELQTTLGFNYFHNEYGKNRFPPEELGLFFDGPDDQDNGLYSY--LGRFKGDKG 481
TF LP+E +L+TTLGFNYF NEY KNRFPPEEL LF++ D GLYS+ GR+ G K
Sbjct: 445 HTFLLPKEIDLKTTLGFNYFTNEYSKNRFPPEELSLFYNDASHDQGLYSHSKRGYSGTGS 504

45
Query: 482 LLPQKSTIVQAGSQYFNTFYFDAALKKDIYRLNYSTNAINYRFGGEYTGYYGSENEFKR 541
LLPQ+S I+QP+G Q F T YFD AL K IY LNYS N +Y F GEY GY
Sbjct: 505 LLPQRSVILQPSGKQKFKTVYFDLTSKGIYHLNYSVNETHYAFNGEYVGY----- 555

50
Query: 542 AFGENSPAYKEHCDPSCGLYEPVLKKGKRRNNHVSISADFGDYFMPFAGYSRTHRMP 601
EN+ + EP+L K G K+A NHS ++SA+ DYFMPF YSRTHRMP
Sbjct: 556 ---ENTAGQQ-----INEPILHKS GHKAFNHSATLSAELSDYFMPFFTYSRTHRMP 604

55
Query: 602 NIQEMYFSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDILGLKLVGYRSRIDNYI 661
NIQEM+FSQ+ ++GV+TALKPE+++T+Q GFNTYKKGL QDD+LG+KLVGYRS I NYI
Sbjct: 605 NIQEMFFSQVSNAGVNTALKPEQSDTYQLGFNTYKKGLFTQDDVLGVKLVGYRSFIKNIYI 664

60
Query: 662 HNVYGKWWDLNGDIPSWVGSTGLAYTIRHRNFKDKVHKHGFLELNYDYGRFFTNLSYAY 721
HNVYG WW +P+W S G YTI H+N+K V K G ELE+NYD GRFF N+SYAY
Sbjct: 665 HNVYGVWW--RDGMPTWAESNGFKYTIHQNYKPIVKKSGVELEINYDMGRFFANVSAY 722

65
Query: 722 QKSTQPTNFS DASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKLTLLGGA 781
Q++ OPTN++DAS PNNAS+ED LKQGYGLSRVS LP+DYGRLE+GTRW KLTIG A
Sbjct: 723 QRTNQPTNYADASPRPNNASQEDILKQGYGLSRVSMLEPKDYGRLELGTWFDQKLTLLGLA 782

70
Query: 782 MRYFGKSIRATAEERYIDGTNGGNTSNVRQLGKRSIKQTETLARQPLIFDYAAAYEPKKN 841
RY+GKS RAT EE YI+G+ + +R+ ++K+TE + +QP+I D + +YEP K+
Sbjct: 783 ARYYGKSKRATIEEEYINGSR-FKKNTLRRENYAVKKTEDIKKQPIILDHVSYPEIKD 841

Query: 842 LIFRAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDKDEDTVCTNADKTLNKGKYGTS 901
LI +AEV+NL D+RY+DPLDAGNDAA+QRYYS + + C D + C GG+
Sbjct: 842 LIIKAEVQNLLDKRYVDPLDAGNDAASQRYYSL-----NNSIECAQDSSAC----GGSD 892

Query: 902 KSVLTNFARGRTFLMTSYKF 922
K+VL NFARGRT+++++YKF
Sbjct: 893 KTVLYNFARGRTYILSLNYKF 913

The underlined motif in the gonococcal protein (also present in the meningococcal protein) is predicted to be an ATP/GTP-binding site motif A (P-loop), and the analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

5 Example 104

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 885>

```

1  ATGAACCTGA  TTTCACGTTA  CATCATCCGT  CAAATGGCGG  TTATGGCGGT
51  TTACGCGCTC  CTTGCCTTCC  TCGCTTTGTA  CAGCTTTTTT  GAAATCCTGT
101 ACGAAACCGG  CAACCTCGGC  AAAGGCAGTT  ACGGCATATG  GGAATGCTG
151 GGCTACACCG  CCTCAAAAT  GCCGCCCCGC  GCCTACGAAC  TGATTCCCCT
201 CGCCGTCCTT  ATCGGCGGAC  TGGTCTCCCT  CAGCCAGCTT  GCCGCCGGCA
251 GCGAACTGAC  CGTCATCAAA  GCCAGCGGCA  TGAGCACCAA  AAAGCTGCTG
301 TTGATTCTGT  CGCAGTTCGG  TTTTATTTTT  GCTATTGCCA  CCGTCGCGCT
351 CGGCGAATGG  GTTGCGCCCA  CACTGAGCCA  AAAAGCCGAA  AACATCAAAG
15  401 CCGCCGCCAT  CAACGGCAAA  ATCAGCACCG  GCAATACCGG  CCTTGGCTG
451 AAAGAAAAAA  ACAGCGTGAT  CAATGTGCGC  GAAATGTTGC  CCGACCAT..

```

This corresponds to the amino acid sequence <SEQ ID 886; ORF112>:

```

1  MNLISRYIIR  QMAVMAVYAL  LAFLALYSFF  EILYETGNLG  KGSYGIWEML
51  GYTALKMPAR  AYELIPLAVL  IGGLVLSLSQ  AAGSELTVIK  ASGMSTKKLL
20  101 LILSQGFIF  AIATVALGEW  VAPTLSQKAE  NIKAAAINGK  ISTGNTGLWL
151 KEKNSVINVR  EMLPDH...

```

Further work revealed further partial nucleotide sequence <SEQ ID 887>:

```

1  ATGAACCTGA  TTTCACGTTA  CATCATCCGT  CAAATGGCGG  TTATGGCGGT
51  TTACGCGCTC  CTTGCCTTCC  TCGCTTTGTA  CAGCTTTTTT  GAAATCCTGT
25  101 ACGAAACCGG  CAACCTCGGC  AAAGGCAGTT  ACGGCATATG  GGAATGCTG
151 gGCTACACCG  CCTCAAAAT  GCCGCCCCGC  GCCTACGAAC  TGATTCCCCT
201 CGCCGTCCTT  ATCGGCGGAC  TGGTCTCCCT  CAGCCAGCTT  GCCGCCGGCA
251 GCGAACTGAC  CGTCATCAAA  GCCAGCGGCA  TGAGCACCAA  AAAGCTGCTG
301 TTGATTCTGT  CGCAGTTCGG  TTTTATTTTT  GCTATTGCCA  CCGTCGCGCT
30  351 CGGCGAATGG  GTTGCGCCCA  CACTGAGCCA  AAAAGCCGAA  AACATCAAAG
401 CCGCCGCCAT  CAACGGCAAA  ATCAGCACCG  GCAATACCGG  CCTTGGCTG
451 AAAGAAAAAA  ACAGCTTKAT  CAATGTGCGC  GAAATGTTGC  CCGACCATAC
501 GCTTTTGGGC  ATCAAAATTT  GGGCGCGCAA  CGATAAAAAC  GAATTGGCAG
551 AGGCAGTGA  AGCCGATTCC  GCCGTTTGA  ACAGCGACGG  CAGTTGGCAG
35  601 TTGAAAAACA  TCCGCCGCG  CACGCTTGGC  GAAGACAAAG  TCGAGGTCTC
651 TATTGCGGCT  GAAGAAAAC  GGCCGATTTC  CGTCAAACGC  AACCTGATGG
701 ACGTATTGCT  CGTCAAACCC  GACCAAATGT  CCGTCGGCGA  ACTGACCACC
751 TACATCCGCC  ACCTCCAAA  CAACAGCCAA  AACACCCGAA  TCTACGCCAT
801 CGCATGGTGG  CGCAAATTGG  TTTACCCGCG  CGCAGCCTGG  GTGATGGCGC
40  851 TCGTCGCCTT  TGCCTTTACC  CCGCAAACCA  CCCGCCACGG  CAATATGGGC
901 TTAAACTCT  TCGGCGGCAT  CTGTstCGGA  TTGCTGTTC  ACCTTGCCGG
951 ACGGCTCTTT  GGGTTTACCA  GCCAACTCGG...

```

This corresponds to the amino acid sequence <SEQ ID 888; ORF112-1>:

```

1  MNLISRYIIR  QMAVMAVYAL  LAFLALYSFF  EILYETGNLG  KGSYGIWEML
45  51  GYTALKMPAR  AYELIPLAVL  IGGLVLSLSQ  AAGSELTVIK  ASGMSTKKLL
101 LILSQGFIF  AIATVALGEW  VAPTLSQKAE  NIKAAAINGK  ISTGNTGLWL
151 KEKNSXINVR  EMLPDHTLLG  IKIWARNDKN  ELAEAVEADS  AVLNSDGSWQ
201 LKNIRRTLIG  EDKVEVSIAA  EENWPISVKR  NLMDVLLVKP  QDMSVVELTT
251 YIRHLQNSQ  NTRIYAIWW  RKLVPAAAW  VMALVAFAFT  PQTRRHGNGM
50  301 LKLFGGICXG  LLFHLAGRLF  GFTSQL...

```

Computer analysis of this amino acid sequence predicts two transmembrane domains and gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF112 shows 96.4% identity over a 166aa overlap with an ORF (ORF112a) from strain A of *N.*

meningitidis:

		10	20	30	40	50	60
5	orf112.pep	MNLISRYIIROMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR					
	orf112a	MNLISRYIIROMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR					
		10	20	30	40	50	60
10	orf112.pep	AYELIPLAVLIGGLVLSQLAAGSELTVIKASGMSTKKLLILSQFGFIFAIATVALGEW					
	orf112a	AYELMPLAVLIGGLVLSXSQLAAGSELXVIKASGMSTKKLLILSQFGFIFAIATVALGEW					
		70	80	90	100	110	120
15	orf112.pep	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINVREMLPDH					
	orf112a	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSIINVREMLPDHTLLGIKIWARNDKN					
		130	140	150	160		
20	orf112a	ELAEAVEADSAVLNSDGSWQLKNIRRSTLGEDKVEVSIAAEEXWPISVKNRLMDVLLVKP					
		190	200	210	220	230	240

The ORF112a nucleotide sequence <SEQ ID 889> is:

25	1	ATGAACCTGA	TTTCACGTTA	CATCATCCGT	CAAATGGCGG	TTATGGCGGT
	51	TTACGCGCTC	CTTGCCCTTCC	TCGCTTTGTA	CAGCTTTTTT	GAAATCCTGT
	101	ACGAAACCGG	CAACCTCGGC	AAAGGCAGTT	ACGGCATATG	GGAATGNTG
	151	GGNTACACCG	CCCTCAAAAT	GNCCGCCCGC	GCCTACGAAC	TGATGCCCTT
	201	CGCCGTCCTT	ATCGGCGGAC	TGGTCTCTNT	CAGCCAGCTT	GCCGCCGGCA
30	251	GCGAACTGAN	CGTCATCAAA	GCCAGCGGCA	TGAGCACCAA	AAAGCTGCTG
	301	TTGATTCTGT	CGCAGTTCGG	TTTTATTTTT	GCTATTGCCA	CCGTGCGCTT
	351	CGGCGAATGG	GTTGCGCCCA	CACTGAGCCA	AAAAGCCGAA	AACATCAAAG
	401	CCGCGGCCAT	CAACGGCAAA	ATCAGTACCG	GCAATACCGG	CCTTGGGCTG
	451	AAAGAAAAAA	ACAGCATTAT	CAATGTGCGC	GAAATGTTGC	CCGACCATAC
35	501	CCTGCTGGGC	ATTAATAATCT	GGGCCCGCAA	CGATAAAAAC	GAAGTGGCAG
	551	AGGCAGTGGA	AGCCGATTCC	GCCGTTTGA	ACAGCGACGG	CAGTTGGCAG
	601	TTGAAAAACA	TCCGCCGCGC	CACGCTTGGC	GAAGACAAAG	TCGAGGTCTC
	651	TATTGCGGCT	GAAGAAAANT	GGCCGATTTC	CGTCAAACGC	AACCTGATGG
	701	ACGTATTGCT	CGTCAAACCC	GACCAAATGT	CCGTGCGCGA	ACTGACCACC
40	751	TACATCGGCC	ACCTCCAAAN	NNACAGCCAA	AACACCCGAA	TCTAGCCCAT
	801	CGCATGGTGG	CGCAAATTGG	TTTACCCCGC	CGCAGCCTGG	GTGATGGCGC
	851	TCGTGCGCTT	TGCCTTTACC	CCGCAAACCA	CCCGCCACGG	CAATATGGGC
	901	TTAAANTCT	TCGGCGGCAT	CTGTCTCGGA	TTGCTGTTCC	ACCTTGCCCG
	951	NCGGCTCTTC	NGGTTTACCA	GCCAACTCTA	CGGCATCCCG	CCCTTCCTCG
45	1001	NCGGCGCACT	ACCTACCATA	GCCTTCGCCT	TGCTCGCCGT	TTGGCTGATA
	1051	CGCAAACAGG	AAAAACGCTA	A		

This encodes a protein having the amino acid sequence <SEQ ID 890>:

	1	MNLISRYIIR	QMAVMAVYAL	LAFLALYSFF	EILYETGNLG	KGSYGIWEMX
	51	GYTALKMXAR	AYELMPLAVL	IGGLVXSXQL	AAGSELXVIK	ASGMSTKKLL
50	101	LILSQFGFIF	AIATVALGEW	VAPTLSQKAE	NIKAAAINGK	ISTGNTGLWL
	151	KEKNSIINVR	EMLPDHTLLG	IKIWARNDKN	ELAEAVEADS	AVLNSDGSWQ
	201	LKNIRRSTLG	EDKVEVSIAA	EEXWPISVKN	NLMDVLLVKP	DOMSVGELTT
	251	YIRHLQXXSQ	NTRIYAIWW	RKLVPAAAW	VMALVAFAT	PQTRHGNMG
	301	LKXFGGICLG	LLFHLAARLF	XFTSOLYGIP	PFLXGALPTI	AFALLAVWLI
55	351	RKQEK*				

ORF112a and ORF112-1 show 96.3% identity in 326 aa overlap:

	orf112a.pep	MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMXGYTALKMXAR	
	orf112-1	MNLISRYIIRQMAVMAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR	
60	orf112a.pep	AYELMPLAVLIGGLVXSXSQLAAGSELXVIKASGMSTKKLLILSQFGFIFAIATVALGEW	

20 Homology with a predicted ORF from *N.gonorrhoeae*

25	orf112.pep	MNLSRYII RQMAVMVAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR	60
	orf112.ng	MNLSRYII RQMAVMVAVYALLAFLALYSFFEILYETGNLGKGSYGIWEMLGYTALKMPAR	60
30	orf112.pep	AYELIPLAVLIGGLVLSQLAAGSELTVIKASGMSTKKLLLSQFGFIFAIATVALGEW	120
	orf112.ng	AYELMPLAVLIGGLASLSQLAAGSELAVIKASGMSTKKLLLSQFGFIFAI A AVALGEW	120
	orf112.pep	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKNSVINVREMLPDH	166
	orf112.ng	VAPTLSQKAENIKAAAINGKISTGNTGLWLKEKTSIINV RGMLPDHTLLG IKIWARNDKN	180

35	1	ATGAACCTGA	TTTCACGTTA	CATCATCCGC	CAAATGGCGG	TTATGGCGGT
	51	TTACGCGCTC	CTTGCCTTCC	TCGCTTTGTA	CAGCTTTTTT	GAAATCCTGT
	101	ACGAAACCGG	CAACCTCGGC	AAAGGCAGTT	ACGGCATATG	GGAAATGCTG
	151	GGCTACACCG	CCCTCAAAAT	GCCC GCCCGC	GCCTACGAAT	TCATGCCCTT
40	201	GCGCGTCTC	ATCGGCGGAC	TGGCCTCTCT	CAGCCAGCTT	GCCGCCGGCA
	251	GCGAACTGGC	CGTCATCAAA	GCCAGCGGCA	TGAGCACCAA	AAAGCTGCTG
	301	TTGATTCTGT	CTCAGTTCGG	TTTATTTTTT	GCTATTGCCG	CCGTTCGCGT
	351	CGCGGAATGG	GTTGCGCCCA	CGCTGAGCCA	AAAAGCCGAA	AACATCAAag
45	401	cGCGCGCCAT	taacggCAAA	ATCAGCAcgg	cgAATACCGG	CCTTTggcTG
	451	AAAGAAAAAa	ccAGCATTAT	CAATGTGcGc	GGAAATGTTG	CCGACCATAC
	501	GCTTTTGGGC	ATCAAAATTT	GGGCGCGCAA	CGATAAAAAA	GAATTTGGCAG
	551	AGGCAGTAGA	AGCCGATTCC	GCCGTTTTGA	ACAGCGACCG	CAGTTGGCAG
50	601	TTGAAAAACA	TCCGCCGCAG	CATCATGGGT	ACAGACAAAA	TCGAAAAATC
	651	cgCCGCCGCC	GAAGAAACTT	gGCCGATTGC	CGTCAGACGC	AACCTGATGG
	701	ACGTATTGCT	CGTCAAGCCC	GACCAAATGT	CCGTCCGCCA	GCTGACCACC
	751	TACATCCGCC	ACCTCCAAAA	CAACAGCCAA	AACACCCAAA	TCTACGCCAT
55	801	CGCATGGTGG	CGTAAACTCG	TTTACCCCGT	CGCCGATATG	GTCATGGCGC
	851	TCGTTGCCTT	CGCTTTTACG	CCGCAAACCA	CGCGCCACGG	CAATATGGGC
	901	TTAAAACTCT	TCGGCGGCAT	CTGTCTCGGA	TTGCTGTTCC	ACCTTGCCGG
	951	CAGGCTCTTC	GGGTTTACCA	GCCAACTCTA	CGGCACCCCA	CCCTTCCTCG
55	1001	CCGGCGCACT	GCCTACCATA	GCCTTCGCCT	TGCTCGCTGT	TTGGCTGATA
	1051	CGCAAAACAGG	AAAAACGTTG	A		

```

1  MNLISRYIIR QMAVMAYVAL LAFLALYSFF EILYETGNLG KGSYGIWEML
51  GYTALKMPAR AVELMPLAVL IGGLEASLSQL AAGSELAVIK ASGMSTKLL
101  LILSQGFIF AIAAVALGEW VAPTLSQKAE NIKAAAINGK ISTGNTGLWL
151  KEKTSIINRV GMLPDHTLLG IKIWARNDNR ELAEAEVADS AVLNSDGSWQ
201  LKNIRRSIMG TDKIETSAAA EETWPIAVRR NLMDEVLVKP DQMSVGELTT
251  YIRHLQNNSS NTQIYAIAWW RKLVPYVAAW VMALVAFAFT POTTRHGNMG
301  LKLFGGICLG LLFHLAGRLF GFTSQLYGTP PFLAGALEPT AFALLAVWLI

```

351 RKQEKR*

ORF112ng and ORF112-1 show 94.2% identity in 326 aa overlap:

		10	20	30	40	50	60
5	orf112ng	MN	LI	SR	YI	IR	QM
	orf112-1	MN	LI	SR	YI	IR	QM
		10	20	30	40	50	60
10	orf112ng	70	80	90	100	110	120
	orf112-1	70	80	90	100	110	120
15	orf112ng	130	140	150	160	170	180
	orf112-1	130	140	150	160	170	180
20	orf112ng	190	200	210	220	230	240
	orf112-1	190	200	210	220	230	240
25	orf112ng	250	260	270	280	290	300
	orf112-1	250	260	270	280	290	300
30	orf112ng	310	320	330	340	350	
	orf112-1	310	320	330	340	350	

This analysis suggests that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their
 40 epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

It will be appreciated that the invention has been described by means of example only, and that
 modifications may be made whilst remaining within the spirit and scope of the invention.

TABLE I - PCR primers

ORF	Primer	Sequence	Restriction sites
ORF 1	Forward Reverse	CGCGGATCCGCTAGC-GGACACACTTATTTTCGG CCCGCTCGAG-CCAGCGGTAGCCTAATT	BamHI-NheI XhoI
ORF 2	Forward Reverse	GCGGATCCCATATG-TTGATTTTCGGTTTGGG CCCGCTCGAG-GACGGCATAACGGCG	BamHI-NdeI XhoI
ORF 2-1	Forward Reverse	GCGGATCCCATATG-TTGATTTTCGGTTTGGG CCCGCTCGAG-TGATTTACGGACGCGCA	BamHI-NdeI XhoI
ORF 4	Forward Reverse	GCGGATCCCATATG-TGCGGAGGTCAAAAAGAC CCCGCTCGAG-TTGCGCTGCGCCTTC	BamHI-NdeI XhoI
ORF 5	Forward Forward Reverse	GGAATTCCATATGGCCATGG-TGGAAGGCGCACAACC CGGGATCC-ATGGAAGGCGCACAAC CCCGCTCGAG-GACTGTGCAAAAACGG	NdeI-NcoI BamHI XhoI
ORF 6	Forward Reverse	CGCGGATCCCATATG-ACCCGTCAATCTCTGCA CCCGCTCGAG-TGCGCCGAACACTTTC	BamHI-NdeI XhoI
ORF 7	Forward Reverse	CGCGGATCCGCTAGC-GCGCTGCTTTTTGTTC CCCGCTCGAG-TTCAAATATATTTGCGGA	BamHI-NheI XhoI
ORF 8	Forward Reverse	GCGGATCCCATATG-GCTCAACTGCTTCGTAC CCCGCTCGAG-AGCAGGCTTTGGCGC	BamHI-NdeI XhoI
ORF 9	Forward Reverse	CGCGGATCCCATATG-CCGAAGGAAGTCGGAAA CCCGCTCGAG-TTCCGAGGTTTTTCGGG	BamHI-NdeI XhoI
ORF 10	Forward Reverse	GCGGATCCCATATG-GACACAAAAGAAATCCTC CCCGCTCGAG-TAATGGGAAACCTTGTTTT	BamHI-NdeI XhoI
ORF 11	Forward Reverse	GCGGATCCCATATG-GCGGTCAACCTCTACG CCCGCTCGAG-GGAAACGACTTCGCC	BamHI-NdeI XhoI
ORF 13	Forward Reverse	CGCGGATCCCATATG-GCTCTGCTTTCCGCGC CCCGCTCGAG-AGGGTGTGTGATAATAAG	BamHI-NdeI XhoI
ORF 15	Forward Forward Reverse	GGAATTCCATATGGCCATGG-GCGGGACACTGACAG CGGGATCC-TGCGGGACACTGACAGG CCCGCTCGAG-AGGTTGGCCTGTCTATG	NdeI-NcoI BamHI XhoI
ORF 17	Forward	GGAATTCCATATGGCCATGG -TTGCCGGCCTGTTCG	NdeI-NcoI

ORF 18	Forward	CGGGATCC-ATTGCCGGCCTGTTCCG	BamHI
	Reverse	CCCGCTCGAG-AAGCAGGTGTACAGC	XhoI
ORF 19	Forward	GCGGATCCCATATG-ATTTTGCTGCATTGGAT	BamHI-NdeI
	Reverse	CCCGCTCGAG-TCTTCCAATTTCTGAAAGC	XhoI
ORF 19	Forward	GGAATTCATATGGCCATGG -TCGCCAGTGT TTTTACC	NdeI-NcoI
	Forward	CGGGATCC-TTCGCCAGTGT TTTTACCG	BamHI
	Reverse	CCCGCTCGAG-GGTGT TTTTGAAGCTGCC	XhoI
ORF 20	Forward	GGAATTCATATGGCCATGG -TCGGCGCGGGTATG	NdeI-NcoI
	Forward	CGGGATCC-TTCGGCGCGGGTATG	BamHI
	Reverse	CCCGCTCGAG-CGGCGAGCGAGAGCA	XhoI
ORF 22	Forward	GGAATTCATATGGCCATGG-TGATTAAATCAAAAAGGTCT	NdeI-NcoI
	Forward	CGGGATCC-ATGATTAAATCAAAAAGGTCTAAACC	BamHI
	Reverse	CCCGCTCGAG-ATTATGATAGCGGCCC	XhoI
ORF 23	Forward	CGCGGATCCCATATG-GATGTTTCTGTTTCAGAC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTAAACCGATAGGTAAACG	XhoI
ORF 24	Forward	GGAATTCATATGGCCATGG -TGATGCCGGAATGGTG	NdeI-NcoI
	Forward	CGGGATCC-ATGATGCCGGAATGGTG	BamHI
	Reverse	CCCGCTCGAG-TGTCAGCGTGGCGCA	XhoI
ORF 25	Forward	GCGGATCCCATATG-TATCGCAAATGATTGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCGATGGAATAGCCG	XhoI
ORF 26	Forward	GCGGATCCCATATG -CAGCTGATCGACTATTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-GACATCGGCGCGTTTT	XhoI
ORF 27	Forward	GGAATTCATATGGCCATGG-AGACCTATTCTGTTTA	NdeI-NcoI
	Forward	CGGGATCC- CAGACCTATTCTGTTTATTTTAATC	BamHI
	Reverse	CCCGCTCGAG-GGGTTCGATTAAATAACCAT	XhoI
ORF 28	Forward	GGAATTCATATGGCCATGG-ACGGCTGTACGTTGATGT	NdeI-NcoI
	Forward	CGGGATCC-AACGGCTGTACGTTGATG	BamHI
	Reverse	CCCGCTCGAG-TTGTCTAGGAATTGCGG	XhoI
ORF 29	Forward	GCGGATCCCATATG -AACGGTTTGGATGCCCCG	BamHI-NdeI
	Forward	CGCGGATCCGCTAGC-AACGGTTTGGATGCCCCG	BamHI-NheI
	Reverse	CCCGCTCGAG-TTGTCTAAGTTCCTGATATG	XhoI
ORF 32	Forward	CGCGGATCCCATATG-AATACTCCTCCTTTTG	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGTATTTTGTATGCTTTG	XhoI
ORF 33	Forward	GCGGATCCCATATG -ATTGATAGGGATCGTATG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGATCTTTCAAACGGCC	XhoI

ORF 35	Forward	GCGGATCCCATATG-TTCAGAGCTCAGCTT	BamHI-NdeI BamHI-NheI XhoI
	Forward	CGCGGATCCGCTAGC-TTCAGAGCTCAGCTT	
	Reverse	CCCGCTCGAG-AAACAGCCATTTGAGCGA	
ORF 37	Forward	GCGGATCCCATATG-GATGACGTATCGGATTTT	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-ATAGCCCGCTTTCAGG	
ORF 58	Forward	CGCGGATCCGCTAGC-TCCGAACGCGAGTGGAT	BamHI-NheI XhoI
	Reverse	CCCGCTCGAG-AGCATTGTCCAAGGGGAC	
ORF 65	Forward	GGAATTCCATATGGCCATGG -TGCTGTATCTGAATCAAG	NdeI-NcoI BamHI XhoI
	Forward	GCGGATCC-TTGCTGTATCTGAATCAAGG	
	Reverse	CCCGCTCGAG-CCGCATCGGCAGACA	
ORF 66	Forward	GCGGATCCCATATG-TACGCATTTACCGCCG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TGGATTTTGACAGATGG	
ORF 72	Forward	CGCGGATCCCATATG- AATGCAGTAAAAATATCTGA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GCCTGAGACCTTTGCAA	
ORF 73	Forward	GCGGATCCCATATG-AGATTTTTCGGTATCGG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTCATCTTTTTCATGTTTCG	
ORF 75	Forward	GCGGATCCCATATG- TCTGTCTTCAAACGGC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTGTGTTTTGCAAGACAG	
ORF 76	Forward	GATCAGCTAGCCATATG-AAACAGAAAAAACCGC	NheI-NdeI BamHI
	Reverse	GCGGATCC-TTACGGTTTGACACCGTT	
ORF 79	Forward	CGCGGATCCCATATG-GTTTCCGCCGCCG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GTGCTGATGCGCTTCG	
ORF 83	Forward	GCGGATCCCATATG-AAAACCTGCTGCTGC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GCCGCCCTTGCGGC	
ORF 84	Forward	GCGGATCCCATATG-GCAGAGATCTGTTTG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GTTTGCCGATCCGACCA	
ORF 85	Forward	CGCGGATCCCATATG- GCGGTTTGGGGCGGA	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TCGGCGCGCGGGC	
ORF 89	Forward	GGAATTCCATATGGCCATGG-CCATACCTTCTTATCA	NdeI-NcoI BamHI XhoI
	Forward	GCGGATCC-GCCATACCTTCTTATCAGAG	
	Reverse	CCCGCTCGAG-TTTTTTGCGATTAGAAAAAGC	
ORF 97	Forward	GCGGATCCCATATG-CATCCTGCCAGCGAAC	BamHI-NdeI

	Reverse	CCCGCTCGAG-TTCGCCTACGGTTTTTTG	XhoI
ORF 98	Forward	GCGGATCCCATATG-ACGGTAACTGCGG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGTTGTTCTGGGCAAATC	XhoI
ORF 100	Forward	GCGGATCCCATATG-TCGGGCATTTACACCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-ACGGGTTTCGGCGGAA	XhoI
ORF 101	Forward	GCGGATCCCATATG-ATTTATCAAAGAAACCTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTTCCGCCTTTCAATGT	XhoI
ORF 102	Forward	GCGGATCCCATATG-GCAGGGCTGTTTTACC	BamHI-NdeI
	Reverse	CCCGCTCGAG-AAACGGTTTGAACACGAC	XhoI
ORF 103	Forward	GCGGATCCCATATG-AACCACGACATCAC	BamHI-NdeI
	Reverse	CCCGCTCGAG-CAGCCACAGGACGGC	XhoI
ORF 104	Forward	GCGGATCCCATATG-ACGTGGGGAACGC	BamHI-NdeI
	Reverse	CCCGCTCGAG-GCGGCGTTTGAACGGC	XhoI
ORF 105	Forward	GCGGATCCCATATG-ACCAAATTTCAAACCCCTC	BamHI-NdeI
	Reverse	CCCGCTCGAG-TAAACGAATGCCGTCCAG	XhoI
ORF 106	Forward	GCGGATCCCATATG-AGGATAACCGACGGCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-TTGTTCCCGATGATGT	XhoI
ORF 109	Forward	GCGGATCCCATATG-GAAGATTTATATATAATACTCG	BamHI-NdeI
	Reverse	CCCGCTCGAG-ATCAGCTTCGAACCGAAG	XhoI
ORF110	Forward	AAAGAATTC-ATGAGTAAATCCCGTAGATCTCCC	EcoRI
	Reverse	AAACTGCAG-GGAAAACCACATCCGCACTCTGCC	PstI
ORF111	Forward	AAAGAATTC-GCACC GAAAAGGCAAAAACCGCA	EcoRI
	Reverse	AAACTGCAG-TCTGCGCTTTTCGGGCAGGGTGG	PstI
ORF113	Forward	AAAGAATTC-ATGAACAAAACCTCTATCGTGTGATTTTCAACCG	EcoRI
	Reverse	AAACTGCAG-TTACGAATGCCTGCTTGCTCGACCGTACTG	PstI
ORF115	Forward	AAAGAATTC-TTGCTTGTGCAAACAGAAAAGACGG	EcoRI
	Reverse	AAAAAAGTCGAC-CTATTTTTAGGGGCTTTTGC TTGTTGAAAAGCCTGCC	SalI
ORF119	Forward	AAAGAATTC-TACAACATGTATCAGGAAAACCAATACCG	EcoRI
	Reverse	AAACTGCAG-TTATGAAAACAGGCGCAGGGCGGTTTGGC	PstI
ORF120	Forward	AAAGAATTC-GCAAGGCTACCCCAATCCGCCGTG	EcoRI
	Reverse	AAACTGCAG-CGTTTGGCTGCCTGGCCGTTGAT	PstI
ORF121	Forward	AAAGAATTC-GCCTTGGTCTGGCTGGTTTTTCGC	EcoRI
	Reverse	AAACTGCAG-TCATCCGCCACCCACCTCGGCCATCCATC	PstI

ORF122	Forward Reverse	AAAAAAGTCGAC-ATGCTTACCGCGCAAGCAGTTC TCC AAACTGCAG-TCAGGAACACAAACGATGACGAATATCCGTATC	Sall PstI
ORF125	Forward Reverse	AAAGAATTC-GCGCTGTTTTTTCGGCGGCGGTAT AAACTGCAG-CGCCGTTTCAAGACGAAAAAGTCG	EcoRI PstI
ORF126	Forward Reverse	AAAGAATTC-GCGGAAACGGTCAAG AAACTGCAG-TTAATCTGTCTTCCGATATAC	EcoRI PstI
ORF127	Forward Reverse	AAAGAATTC-ATGACTGATAATCGGGGGTTTACG AAAAAAGTCGAC-CTTAAGTAACTGCAGTCCTTATC	EcoRI Sall
ORF128	Forward Reverse	AAAGAATTC-ATGCAAGCTGTCCGTACAGGCC AAACTGCAG-CTATTGCAATGCGCCGCCGCGGAATGTTTGAGCAGGCG	EcoRI PstI
ORF129	Forward Reverse	AAAGAATTC-ATGGATTTTCGTTTTGACATTATTTACGAATACCG AAACTGCAG-TTATTTTTTGATGAAATTTTGGGGCGG	EcoRI PstI
ORF130	Forward Reverse	AAAGAATTC-GCAGTACTTGCCAT TCTCGGTGCG AAACTGCAG-CTCCGGATCGTCTGTAAACGCATT	EcoRI PstI
ORF 131	Forward Reverse	GCGGATCCCATATG-GAAATTCGGGCAATAAAAT CCCGCTCGAG-CCAGCGGACGCGTTC	BamHI-NdeI XhoI
ORF 132	Forward Reverse	GCGGATCCCATATG-AAAGAAGCGGGGTTTG CCCGCTCGAG-CCAATCTGCCAGCCGT	BamHI-NdeI XhoI
ORF 133	Forward Reverse	CGCGGATCCCATATG-GAAGATGCAGGGCGCG CCCGCTCGAG-AAACTGTAGCTCATCGT	BamHI-NdeI XhoI
ORF 134	Forward Reverse	GCGGATCCCATATG-TCTGTGCAAGCAGTATTG CCCGCTCGAG-ATCCTGTGCCAATGCG	BamHI-NdeI XhoI
ORF 135	Forward Reverse	GCGGATCCCATATG-CCGTCTGAAAAAGCTTT CCCGCTCGAG-AAATACCGCTGAGGATG	BamHI-NdeI XhoI
ORF 136	Forward Reverse	CGCGGATCCGCTAGC-ATGAAGCGGCGTATAGCC CCCGCTCGAG-TTCCGAATATTTGGAACTTT	BamHI-NheI XhoI
ORF 137	Forward Reverse	CGCGGATCCCATATG-GGCACGGCGGGAAATA CCCGCTCGAG-ATAACGGTATGCCGCC	BamHI-NdeI XhoI
ORF 138	Forward Reverse	GCGGATCCCATATG-TTTCGTTTACAATTCAGGC CCCGCTCGAG-CGGCGTTTTATAGCGG	BamHI-NdeI XhoI
ORF 139	Forward Reverse	GCGGATCCCATATG-GCTTTTTTGCGGTAATG CCCGCTCGAG-TAACGTTTCCGTGCGTTT	BamHI-NdeI XhoI

ORF 140	Forward	GCGGATCCCATATG-TTGCCACAGGCAGC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-GACGATGGCAAACAGC	
ORF 141	Forward	GCGGATCCCATATG-CCGTCTGAAGCAGTCT	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-ATCTGTTGTTTTTAAATATT	
ORF 142	Forward	GCGGATCCCATATG-GATAATTCTGGTAGTGAAG	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-AAACGTATAGCCTACCT	
ORF 143	Forward	GCGGATCCCATATG-GATACCGCTTTGAACCT	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-AATGGCTCCGCAATATG	
ORF 144	Forward	GCGGATCCCATATG-ACCTTTTACAACGTTTGC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-AGATTGTTGTTGTTTTTTCG	
ORF 147	Forward	GCGGATCCCATATG-TCTGTCTTTCAAACGGC	BamHI-NdeI XhoI
	Reverse	CCCGCTCGAG-TTGTTTTTGCAAGACAG	

NB:

- restriction sites are underlined
- for ORFs 110-130, where the ORF itself carries an *EcoRI* site (eg. ORF122), a *SalI* site was used in the forward primer instead. Similarly, where the ORF carries a *PstI* site (eg. ORFs 115 and 127), a *SalI* site was used in the reverse primer.

5

TABLE II – Summary of cloning, expression and purification

ORF	PCR/cloning	His-fusion expression	GST-fusion expression	Purification
orf 1	+	+	+	His-fusion
orf 2	+	+	+	GST-fusion
orf 2.1	+	n.d.	+	GST-fusion
orf 4	+	+	+	His-fusion
orf 5	+	n.d.	+	GST-fusion
orf 6	+	+	+	GST-fusion
orf 7	+	+	+	GST-fusion
orf 8	+	n.d.	n.d.	
orf 9	+	+	+	GST-fusion
orf 10	+	n.d.	n.d.	
orf 11	+	n.d.	n.d.	
orf 13	+	n.d.	+	GST-fusion
orf 15	+	+	+	GST-fusion
orf 17	+	n.d.	n.d.	
orf 18	+	n.d.	n.d.	
orf 19	+	n.d.	n.d.	
orf 20	+	n.d.	n.d.	
orf 22	+	+	+	GST-fusion
orf 23	+	+	+	His-fusion
orf 24	+	n.d.	n.d.	
orf 25	+	+	+	His-fusion
orf 26	+	n.d.	n.d.	
orf 27	+	+	+	GST-fusion
orf 28	+	+	+	GST-fusion
orf 29	+	n.d.	n.d.	
orf 32	+	+	+	His-fusion
orf 33	+	n.d.	n.d.	
orf 35	+	n.d.	n.d.	
orf 37	+	+	+	GST-fusion
orf 58	+	n.d.	n.d.	
orf 65	+	n.d.	n.d.	
orf 66	+	n.d.	n.d.	
orf 72	+	+	n.d.	His-fusion
orf 73	+	n.d.	+	n.d.
orf 75	+	n.d.	n.d.	
orf 76	+	+	n.d.	His-fusion
orf 79	+	+	n.d.	His-fusion
orf 83	+	n.d.	+	n.d.
orf 84	+	n.d.	n.d.	

orf 85	+	n.d.	+	GST-fusion
orf 89	+	n.d.	+	GST-fusion
orf 97	+	+	+	GST-fusion
orf 98	+	n.d.	n.d.	
orf 100	+	n.d.	n.d.	
orf 101	+	n.d.	n.d.	
orf 102	+	n.d.	n.d.	
orf 103	+	n.d.	n.d.	
orf 104	+	n.d.	n.d.	
orf 105	+	n.d.	n.d.	
orf 106	+	+	+	His-fusion
orf 109	+	n.d.	n.d.	
orf 110	+	n.d.	n.d.	
orf 111	+	+	n.d.	His-fusion
orf 113	+	+	n.d.	His-fusion
orf 115	n.d.	n.d.	n.d.	
orf 119	+	+	n.d.	His-fusion
orf 120	+	+	n.d.	His-fusion
orf 121	+	n.d.	n.d.	
orf 122	+	+	n.d.	His-fusion
orf 125	+	+	n.d.	His-fusion
orf 126	+	+	n.d.	His-fusion
orf 127	+	+	n.d.	His-fusion
orf 128	+	n.d.	n.d.	
orf 129	+	+	n.d.	His-fusion
orf 130	+	n.d.	n.d.	
orf 131	+	+	+	n.d.
orf 132	+	+	+	His-fusion
orf 133	+	n.d.	+	GST-fusion
orf 134	+	n.d.	n.d.	
orf 135	+	n.d.	n.d.	
orf 136	+	n.d.	n.d.	
orf 137	+	n.d.	+	GST-fusion
orf 138	+	n.d.	+	GST-fusion
orf 139	+	n.d.	n.d.	
orf 140	+	n.d.	n.d.	
orf 141	+	n.d.	n.d.	
orf 142	+	n.d.	n.d.	
orf 143	+	n.d.	n.d.	
orf 144	+	n.d.	+	n.d.
orf 147	+	n.d.	n.d.	

CLAIMS

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, and 8.
2. A nucleic acid molecule which encodes a protein according to claim 1.
- 5 3. A nucleic acid molecule according to claim 2, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, and 7.
4. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102,
10 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302,
15 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502,
20 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702,
25 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..
- 30 5. A protein having 50% or greater sequence identity to a protein according to claim 4.

6. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, 378, 380, 382, 384, 386, 388, 390, 392, 394, 396, 398, 400, 402, 404, 406, 408, 410, 412, 414, 416, 418, 420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446, 448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478, 480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510, 512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542, 544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574, 576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606, 608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638, 640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670, 672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702, 704, 706, 708, 710, 712, 714, 716, 718, 720, 722, 724, 726, 728, 730, 732, 734, 736, 738, 740, 742, 744, 746, 748, 750, 752, 754, 756, 758, 760, 762, 764, 766, 768, 770, 772, 774, 776, 778, 780, 782, 784, 786, 788, 790, 792, 794, 796, 798, 800, 802, 804, 806, 808, 810, 812, 814, 816, 818, 820, 822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, 866, 868, 870, 872, 874, 876, 878, 880, 882, 884, 886, 888, 890, & 892..
7. An antibody which binds to a protein according to any one of claims 4 to 6.
8. A nucleic acid molecule which encodes a protein according to any one of claims 4 to 6.
9. A nucleic acid molecule according to claim 8, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289,

291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329,
331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369,
371, 373, 375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409,
411, 413, 415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449,
5 451, 453, 455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489,
491, 493, 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529,
531, 533, 535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569,
571, 573, 575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609,
611, 613, 615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649,
10 651, 653, 655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689,
691, 693, 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729,
731, 733, 735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769,
771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809,
811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849,
15 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889,
& 891..

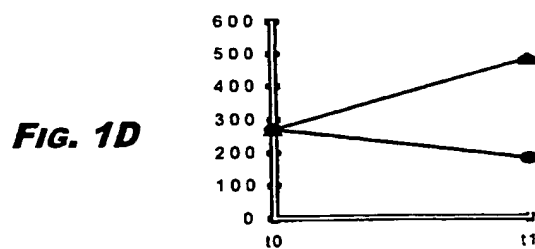
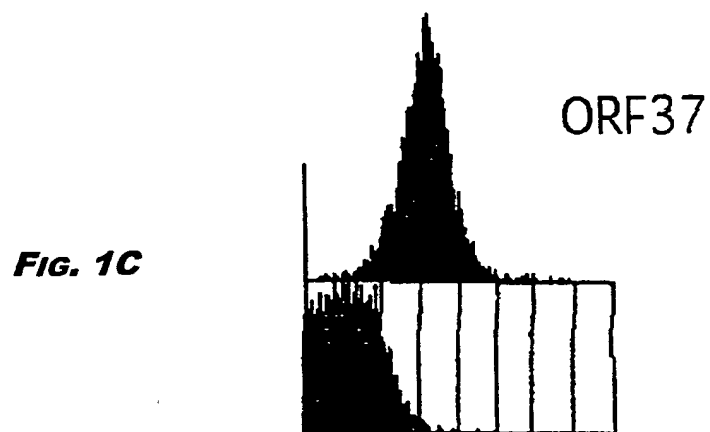
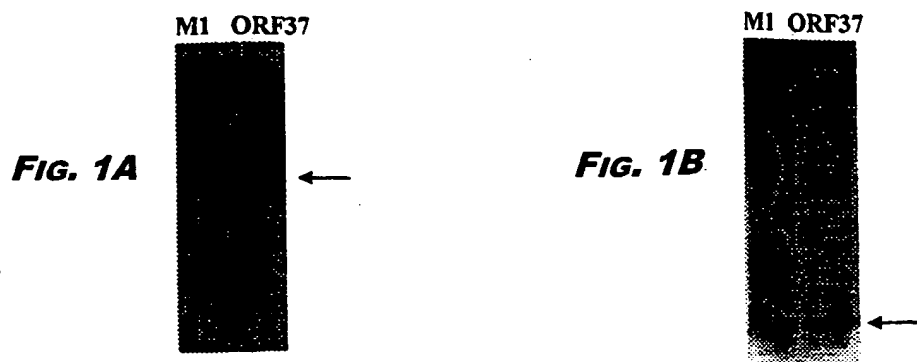
10. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the
group consisting of SEQ IDs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39,
41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91,
20 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133,
135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173,
175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213,
215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253,
255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293,
25 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333,
335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373,
375, 377, 379, 381, 383, 385, 387, 389, 391, 393, 395, 397, 399, 401, 403, 405, 407, 409, 411, 413,
415, 417, 419, 421, 423, 425, 427, 429, 431, 433, 435, 437, 439, 441, 443, 445, 447, 449, 451, 453,
455, 457, 459, 461, 463, 465, 467, 469, 471, 473, 475, 477, 479, 481, 483, 485, 487, 489, 491, 493,
30 495, 497, 499, 501, 503, 505, 507, 509, 511, 513, 515, 517, 519, 521, 523, 525, 527, 529, 531, 533,
535, 537, 539, 541, 543, 545, 547, 549, 551, 553, 555, 557, 559, 561, 563, 565, 567, 569, 571, 573,
575, 577, 579, 581, 583, 585, 587, 589, 591, 593, 595, 597, 599, 601, 603, 605, 607, 609, 611, 613,
615, 617, 619, 621, 623, 625, 627, 629, 631, 633, 635, 637, 639, 641, 643, 645, 647, 649, 651, 653,
655, 657, 659, 661, 663, 665, 667, 669, 671, 673, 675, 677, 679, 681, 683, 685, 687, 689, 691, 693,
35 695, 697, 699, 701, 703, 705, 707, 709, 711, 713, 715, 717, 719, 721, 723, 725, 727, 729, 731, 733,

735, 737, 739, 741, 743, 745, 747, 749, 751, 753, 755, 757, 759, 761, 763, 765, 767, 769, 771, 773, 775, 777, 779, 781, 783, 785, 787, 789, 791, 793, 795, 797, 799, 801, 803, 805, 807, 809, 811, 813, 815, 817, 819, 821, 823, 825, 827, 829, 831, 833, 835, 837, 839, 841, 843, 845, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865, 867, 869, 871, 873, 875, 877, 879, 881, 883, 885, 887, 889, & 891..

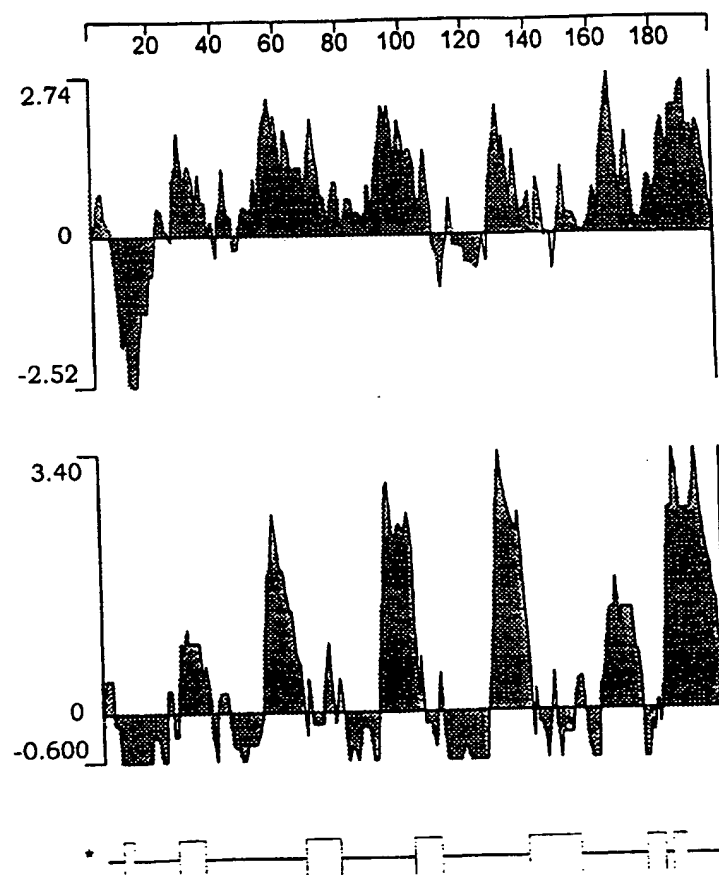
- 5 11. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 8 to 10.
12. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 8-11.
13. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any
10 one of claims 8-12 under high stringency conditions.
14. A composition comprising a protein, a nucleic acid molecule, or an antibody according to any preceding claim.
15. A composition according to claim 14 being a vaccine composition or a diagnostic composition.
- 15 16. A composition according to claim 14 or claim 15 for use as a pharmaceutical.
17. The use of a composition according to claim 14 in the manufacture of a medicament for the treatment or prevention of infection due to Neisserial bacteria.

1/24

FIGURE 1



2/24

FIG 1E

3/24

FIGURE 2

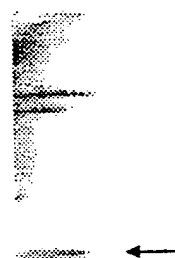
FIG. 2A

M1 ORF5



FIG. 2B

TP



4/24

FIGURE 3

FIG. 3A

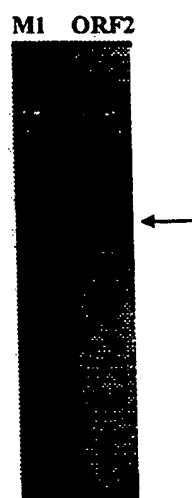


FIG. 3B

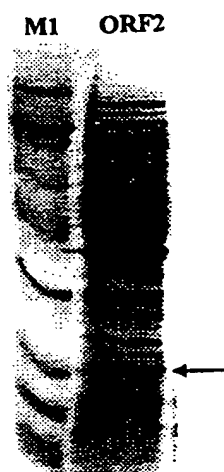


FIG. 3C

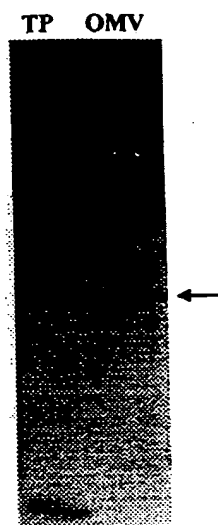
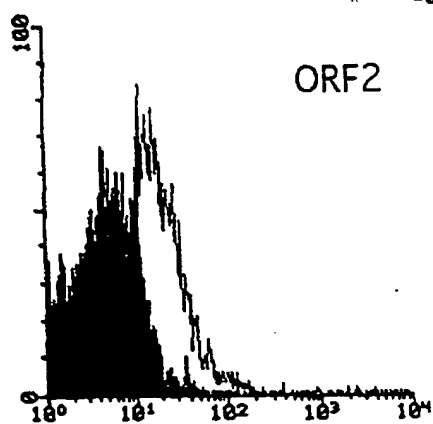


FIG. 3D



5/24

FIGURE 4

FIG. 4A

M1 ORF15



FIG. 4B

M2 ORF15

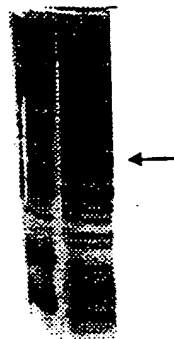
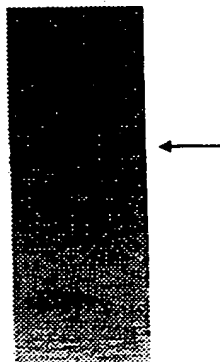


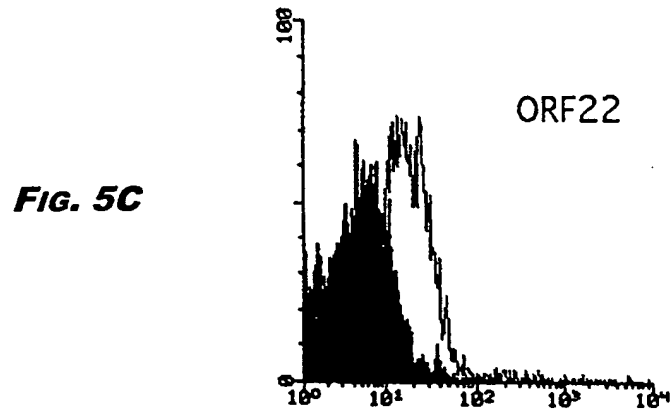
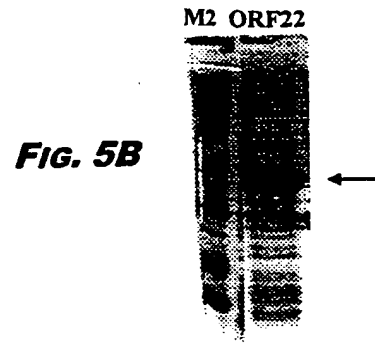
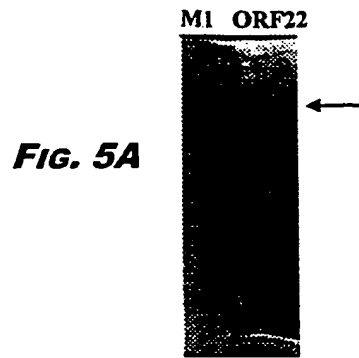
FIG 4C

TP OMV



6/24

FIGURE 5



7/24

FIGURE 6

Fig. 6A

M1 ORF28

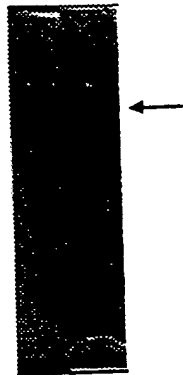


Fig. 6B

M2 ORF28

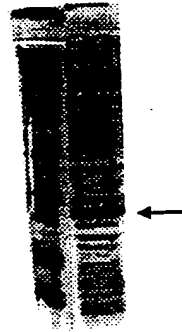


FIGURE 7

Fig. 7A

M1 ORF32

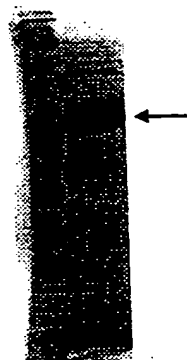
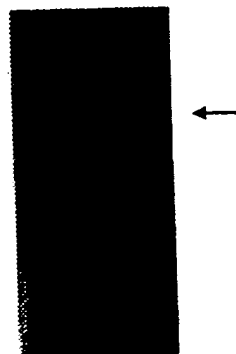


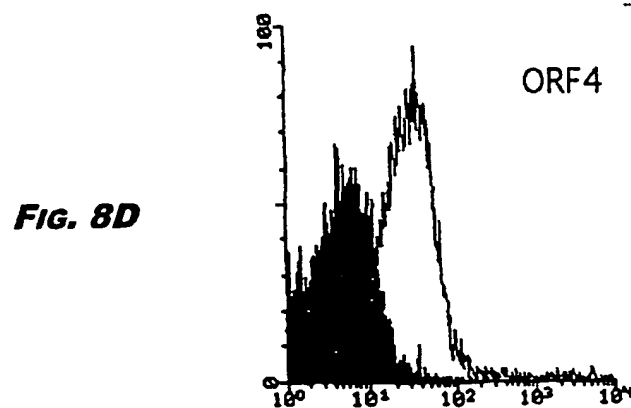
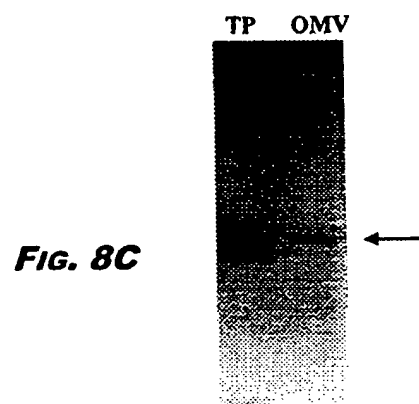
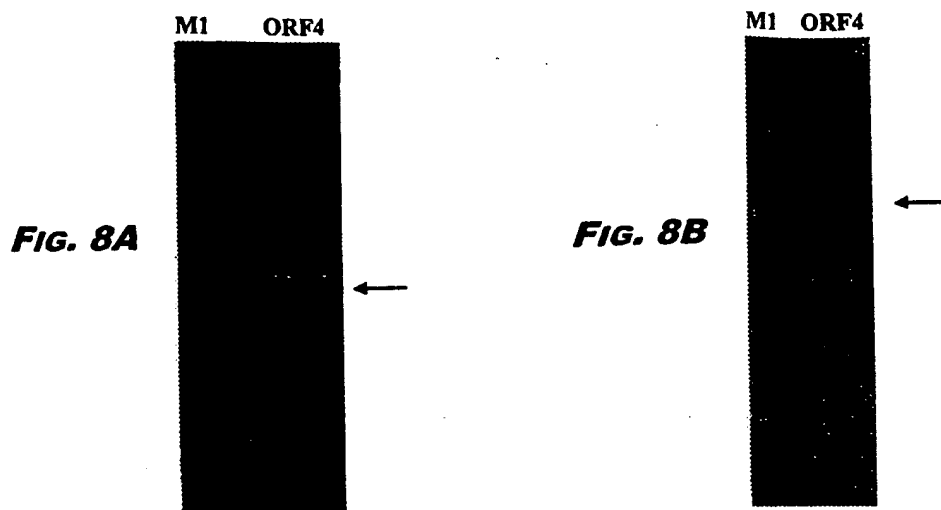
Fig. 7B

M1 ORF32



8/24

FIGURE 8



9/24

FIG. 8E

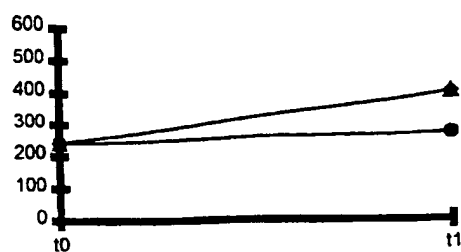
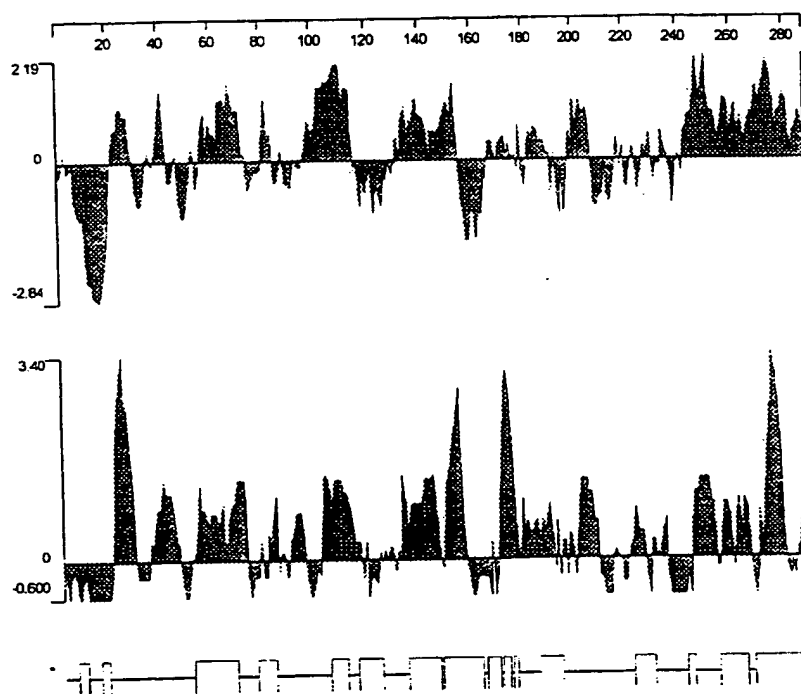
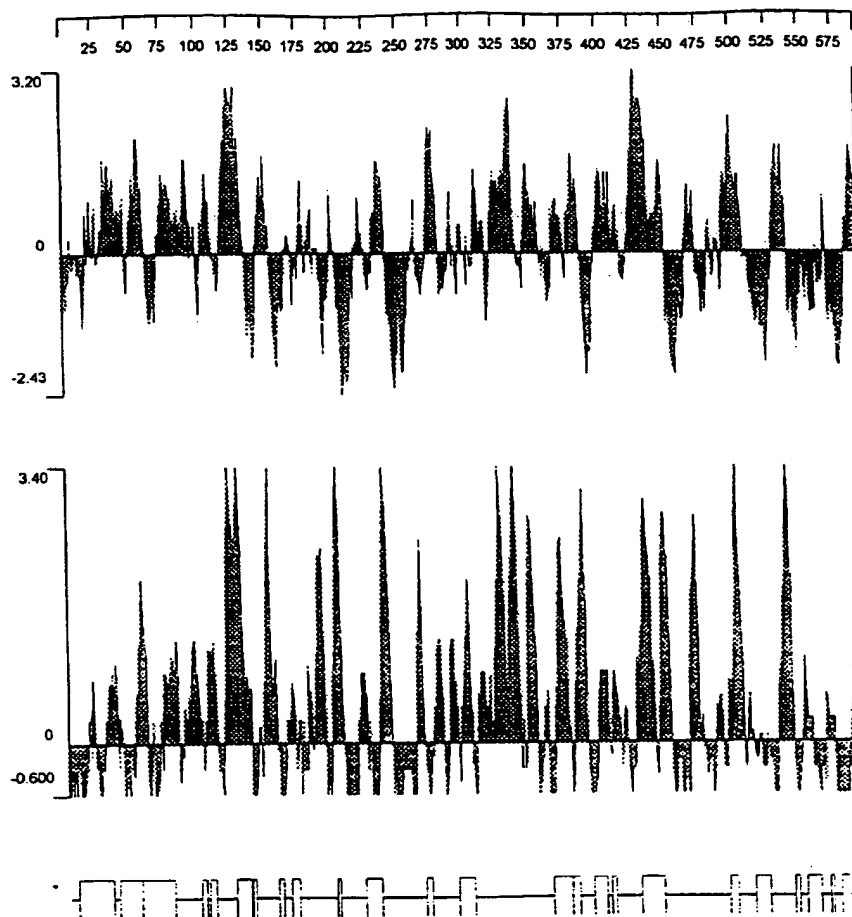


FIG. 8F



10/24

FIGURE 9

11/24

FIGURE 10

FIG. 10A

M1 ORF76

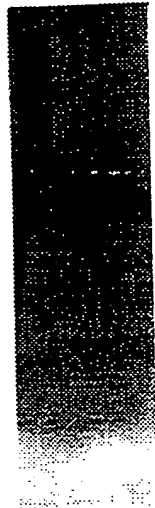
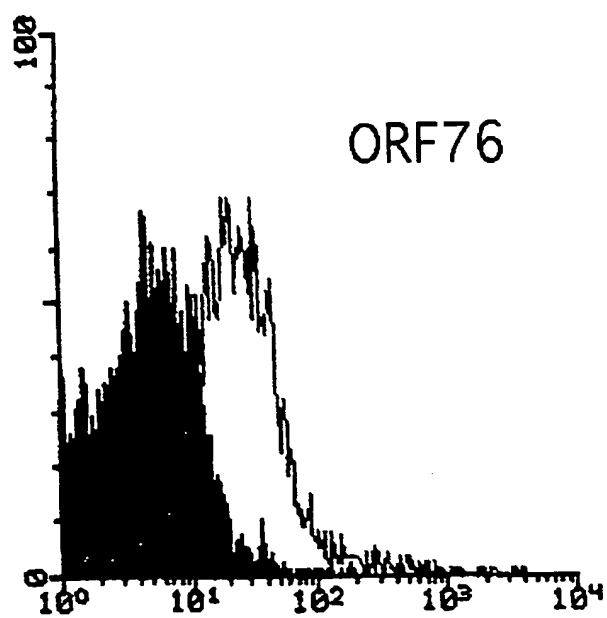


FIG. 10B

TP OMV



FIG. 10C



12/24

FIGURE 11

M1 ORF89



13/24

FIGURE 12

FIG. 12A



FIG. 12B

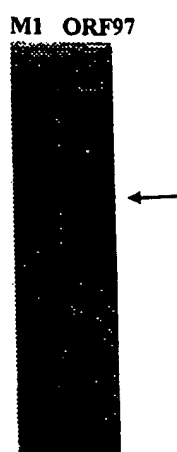


FIG. 12C

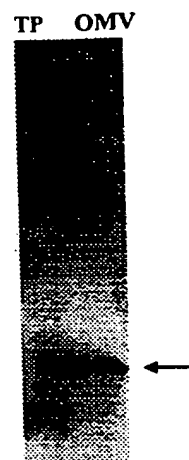
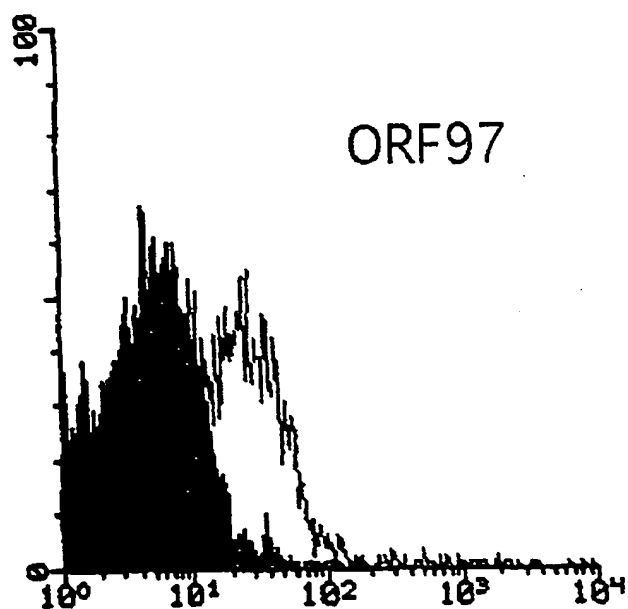
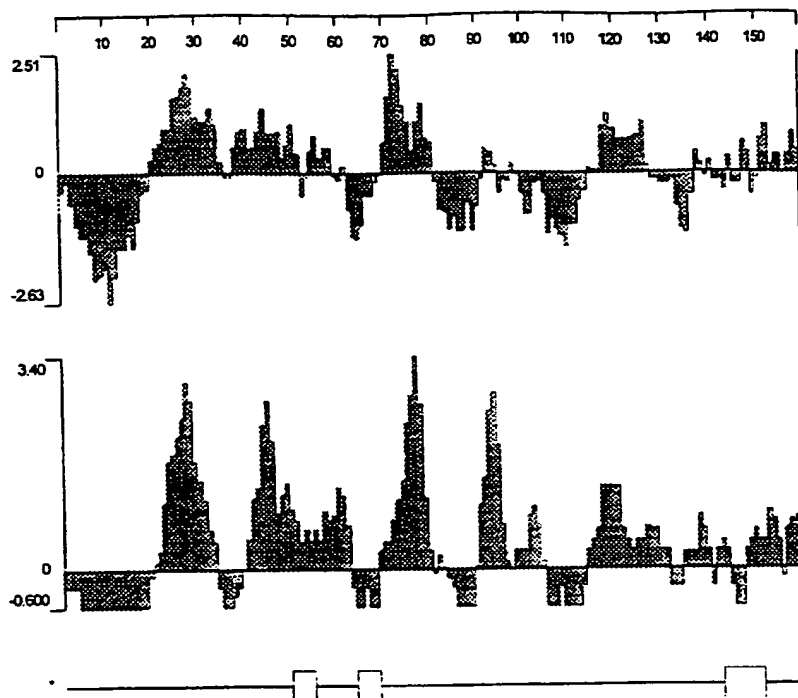


FIG. 12D



14/24

FIG. 12E



15/24

FIGURE 13

Fig. 13A

M1 ORF106

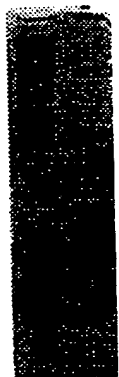
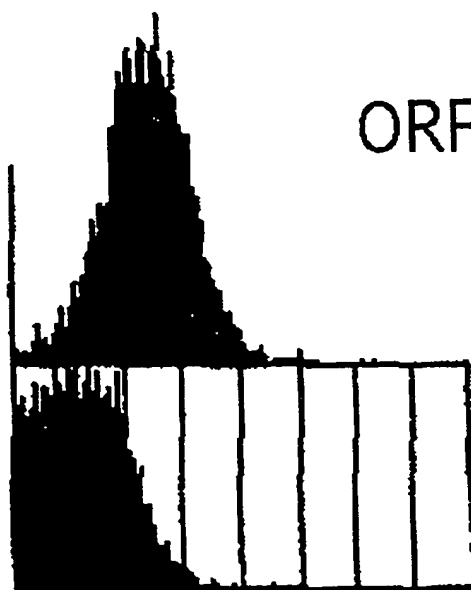


Fig. 13B

M2 ORF106



Fig. 13C



ORF 106

FIGURE 14

FIG. 14A

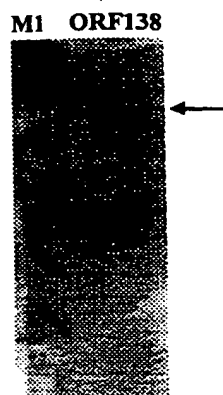
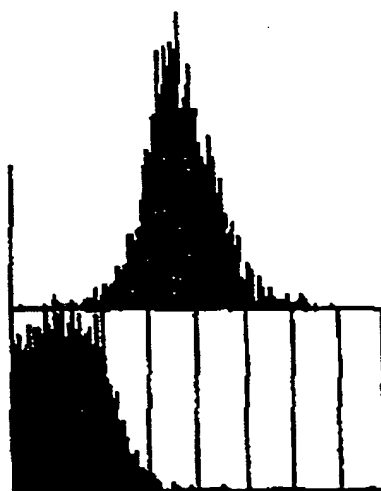


FIG. 14B



17/24

FIGURE 15

FIG. 15A

M1 ORF23



FIG. 15B

M2 ORF23



Fig 15C

TP OMV



18/24

FIGURE 16

FIG. 16A

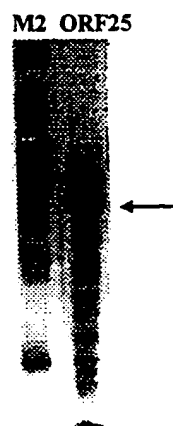


FIG. 16B

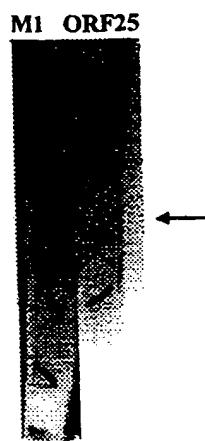


FIG. 16C

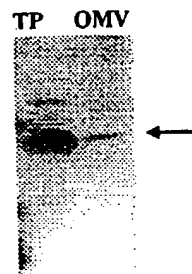
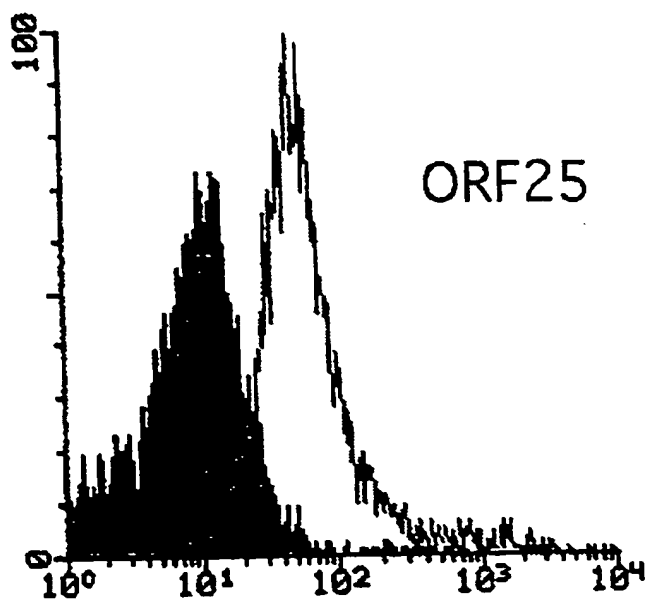


FIG. 16D



19/24

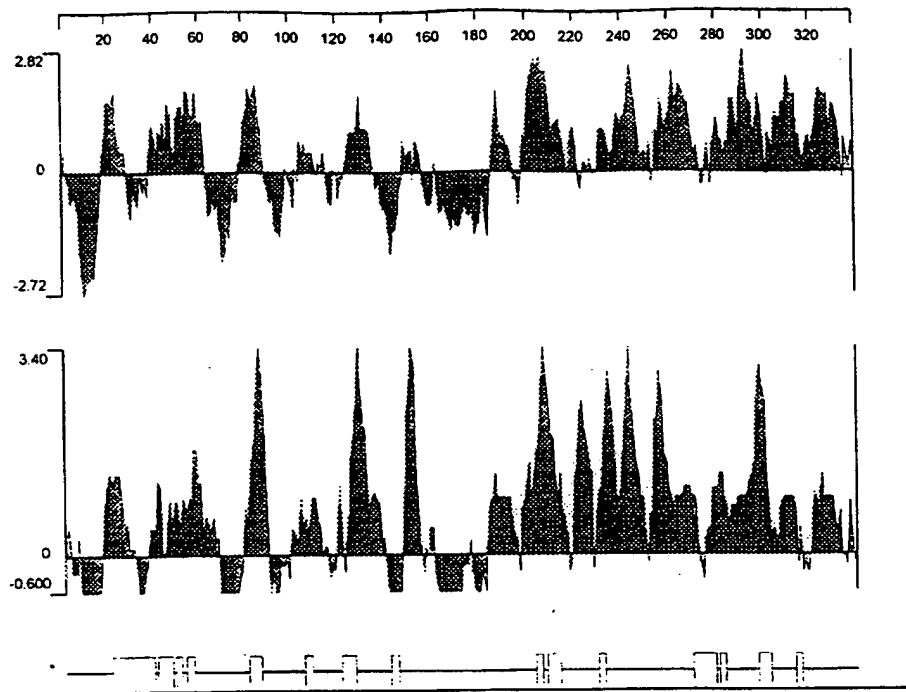
FIG. 16E

FIGURE 17

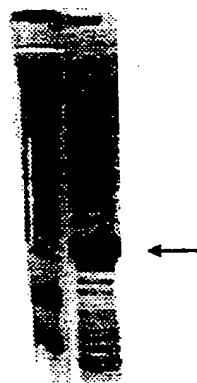
FIG. 17A

M1 ORF27



FIG. 17B

M2 ORF27



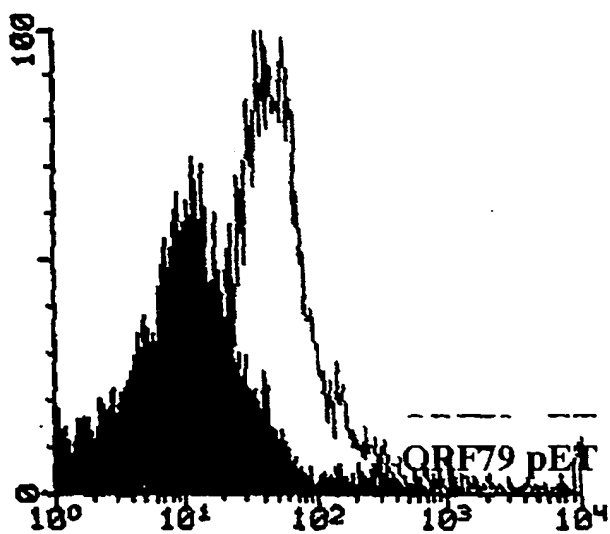
21/24

FIGURE 18

FIG. 18A



FIG. 18B



22/24

FIGURE 19

Fig. 19A

M1 ORF85

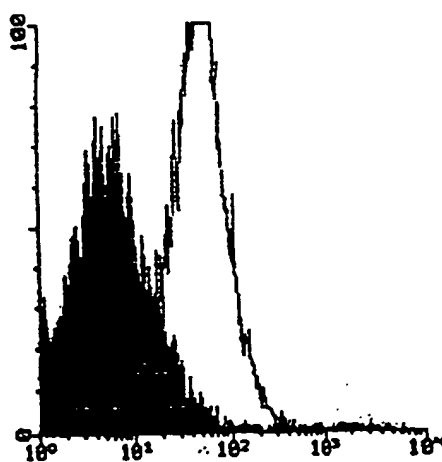


Fig. 19B

TP OMV

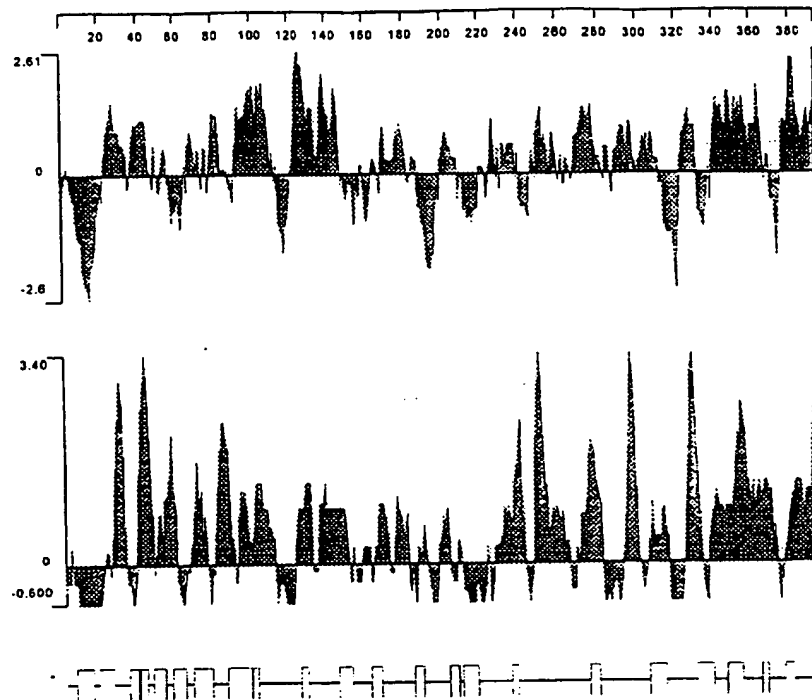


Fig. 19C



23/24

Fig 19D



24/24

FIGURE 20

Fig. 20A

M1 ORF132

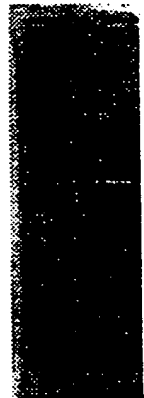
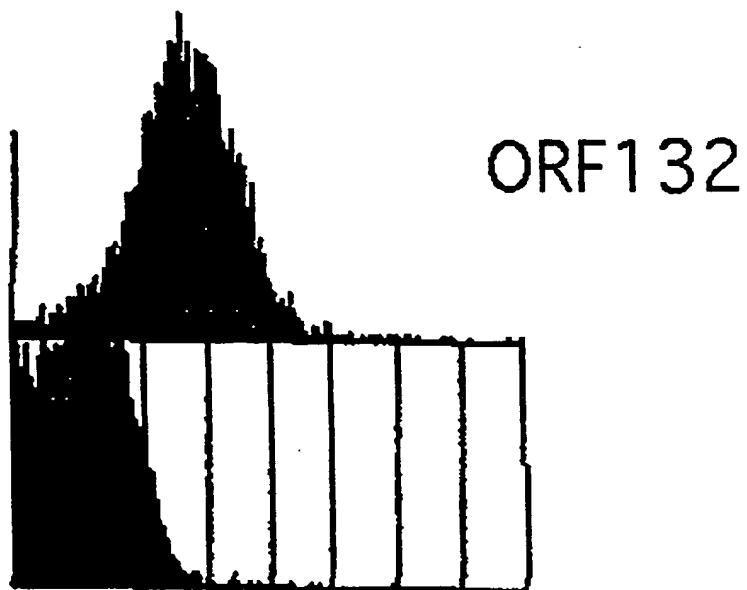


Fig. 20B

M2 ORF132



Fig. 20C



PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/31, C07K 14/22, 16/12, G01N 33/53, A61K 39/095 // (C12N 15/31, C12 R 1:36)		A3	(11) International Publication Number: WO 99/24578 (43) International Publication Date: 20 May 1999 (20.05.99)
(21) International Application Number: PCT/IB98/01665 (22) International Filing Date: 9 October 1998 (09.10.98)		(74) Agent: HALLYBONE, Huw, George; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).	
(30) Priority Data: 9723516.2 6 November 1997 (06.11.97) GB 9724190.5 14 November 1997 (14.11.97) GB 9724386.9 18 November 1997 (18.11.97) GB 9725158.1 27 November 1997 (27.11.97) GB 9726147.3 10 December 1997 (10.12.97) GB 9800759.4 14 January 1998 (14.01.98) GB 9819016.8 1 September 1998 (01.09.98) GB		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(71) Applicant (for all designated States except US): CHIRON S.P.A. [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT).		Published With international search report.	
(72) Inventors; and (75) Inventors/Applicants (for US only): MASNANI, Vega [IT/IT]; Via Pantaneto, 105, I-53100 Siena (IT). RAP- PUOLI, Rino [IT/IT]; Via delle Rocche, 1, Vagliagli, I-53019 Castelnuovo Berardenga (IT). PIZZA, Mariagrazia [IT/IT]; Strada di Montalbuccio, 160, I-53100 Siena (IT). SCARLATO, Vincenzo [IT/IT]; Via Firenze, 3/37, I-53134 Colle Val d'Elsa (IT). GRANDI, Guido [IT/IT]; 9° Strada, 4, I-20090 Segrate (IT).		(88) Date of publication of the international search report: 2 March 2000 (02.03.00)	
(54) Title: NEISSERIAL ANTIGENS			
(57) Abstract <p>The invention provides proteins from <i>Neisseria meningitidis</i> (strains A and B) and from <i>Neisseria gonorrhoeae</i> including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

INTERNATIONAL SEARCH REPORT

International Application No

PC1/IB 98/01665

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/31 C07K14/22 C07K16/12 G01N33/53 A61K39/095
 //(C12N15/31,C12R1:36)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	WO 99 55873 A (SMITHKLINE BEECHAM BIOLOGICALS S.A.) 4 November 1999 (1999-11-04) SEQ ID NOS: 1, 2, 3 and 4 page 82 -page 96 claims 3-5,10,11,17-21 --- -/--	5-8, 10-16

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

1 December 1999

Date of mailing of the international search report

08 December 1999 (08.12.99)

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl
 Fax: (+31-70) 340-3016

Authorized officer

Fuchs, U

6

INTERNATIONAL SEARCH REPORT

International Application No

PC, IB 98/01665

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>ROKBI, B. ET AL.: "Evaluation of Recombinant Transferrin-Binding Protein B Variants from Neisseria meningitidis for Their Ability To Induce Cross-Reactive and Bactericidal Antibodies against a Genetically Diverse Collection of Serogroup B Strains"</p> <p>INFECTION AND IMMUNITY, vol. 65, no. 1, January 1997 (1997-01), pages 55-63, XP002086937</p> <p>abstract</p> <p>page 55, column 1, line 1 -page 56, column 2, line 78</p> <p>page 57, column 1, line 31 -page 61, column 2, line 63</p> <p>page 59; figure 3; table 2</p> <p>---</p>	1-17
A	<p>WO 96 12020 A (OREGON HEALTH SCIENCES UNIVERSITY) 25 April 1996 (1996-04-25)</p> <p>abstract</p> <p>page 32 -page 38; examples 2-4</p> <p>page 42 -page 44; example 9</p> <p>page 45; table 3</p> <p>---</p>	1-17
A	<p>WO 96 31618 A (THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL)</p> <p>10 October 1996 (1996-10-10)</p> <p>abstract</p> <p>page 19, line 21 -page 20, line 13</p> <p>page 20, line 22 -page 21, line 11</p> <p>page 22, line 23 -page 24, line 16</p> <p>-----</p>	1-17

INTERNATIONAL SEARCH REPORT

In. ational application No.
PCT/IB 98/01665

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
((1-3) completely) and ((4-17) partially)
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☒ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1. Claims: ((1-3) completely) and ((4-17) partially)

A protein comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 2,4,6 and 8 or fragments thereof; a protein having 50% or greater sequence identity to said protein; an antibody binding said protein; a nucleic acid encoding said protein; a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOS: 1,3,5 and 7 or fragments thereof; a composition comprising said protein, said nucleic acid or said antibody; the use of said composition;

Invention 2 to 104. Claims (4-17) partially

Idem as subject 1 but limited to the ORFs corresponding to examples 2-104 characterized by SEQ ID NOS: 9-892.
(Invention 2 is limited to SEQ ID NOS: 9-10; Invention 3 is limited to SEQ ID NOS: 11-18; Invention 4 is limited to SEQ ID NOS: 19-28;; Invention 104 is limited to SEQ ID NOS: 885-892).

In view of additional search fees paid, Inventions 5, 26, 55, 77 and 91 have been further searched.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PC1, IB 98/01665

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9955873	A	04-11-1999	NONE
WO 9612020	A	25-04-1996	US 5698438 A 16-12-1997
		AU 705509 B 27-05-1999	
		AU 4007395 A 06-05-1996	
		CA 2203116 A 25-04-1996	
		EP 0793720 A 10-09-1997	
		FI 971634 A 16-06-1997	
		HU 77048 A 02-03-1998	
		JP 10508469 T 25-08-1998	
		NO 971768 A 03-06-1997	
WO 9631618	A	10-10-1996	AU 5537096 A 23-10-1996
			CA 2217522 A 10-10-1996
			EP 0830456 A 25-03-1998
			JP 11503322 T 26-03-1999